Machine-Learning-Driven Haptic Sensor Design

Machine-Learning-Driven Haptic Sensor Design

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Huanbo Sun

aus Jiangsu, China

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	13.06.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Dr. Georg Martius
2. Berichterstatter:	Prof. Dr. Hendrik P. A. Lensch
3. Berichterstatter:	Prof. Dr. Helge Ritter

Abstract

Similar to biological systems, robots may need skin-like sensing ability to perceive interactions in complex, changing, and human-involved environments. Current skin-like sensing technologies are still far behind their biological counterparts when considering resolution, dynamics range, robustness, and surface coverage together. One key challenge is the wiring of sensing elements. During my Ph.D. study, I explore how machine learning can enable the design of a new kind of haptic sensors to deal with such a challenge. On the one hand, I propose super-resolution-oriented tactile skins, reducing the number of physical sensing elements while achieving high spatial accuracy. On the other hand, I explore vision-based haptic sensor designs. In this thesis, I present four types of machine-learning-driven haptic sensors that I designed for coarse and fine robotic applications, varying from large surface (robot limbs) to small surface sensing (robot fingers). Moreover, I propose a super-resolution theory to guide sensor designs at all levels ranging from hardware design (material/structure/transduction), data collection (real/simulated), and signal processing methods (analytical/data-driven).

I investigate two designs for large-scale coarse-resolution sensing, e. g., robotic limbs. **HapDef** sparsely attaches a few strain gauges on a large curved surface internally to measure the deformation over the whole surface. **ERT-DNN** wraps a large surface with a piece of multi-layered conductive fabric, which varies its conductivity upon contacts exerted. I also conceive two approaches for small-scale fine-resolution sensing, e. g., robotic fingertips. **BaroDome** sparsely embeds a few barometers inside a soft elastomer to measure internal pressure changes caused by external contact. **Insight** encloses a high-resolution camera to view a soft shell from within.

Generically, an inverse problem needs to be solved when trying to obtain high-resolution sensing with a few physical sensing elements. I develop machine-learning frameworks suitable for solving this inverse problem. They process various raw sensor data and extract useful haptic information in practice. Machine learning methods rely on data collected by an automated robotic stimulation device or synthesized using finite element methods. I build several physical testbeds and finite element models to collect copious data. I propose machine learning frameworks to combine data from different sources that are good enough to deal with the noise in real data and generalize well from seen to unseen situations.

While developing my prototype sensors, I have faced reoccurring design choices. To help my developments and guide future research, I propose a unified theory with the concept of taxel-value-isolines. It captures the physical effects required for super-resolution, ties them to all parts of the sensor design, and allows us to assess them

quantitatively. The theory offers an explanation about physically achievable accuracies for localizing and quantifying contact based on uncertainties introduced by measurement noise in sensor elements. The theoretical analysis aims to predict the best performance before a physical prototype is built and helps to evaluate the hardware design, data collection, and data processing methods during implementation.

This thesis presents a new perspective on haptic sensor design. Using machine learning to substitute the entire data-processing pipeline, I present several haptic sensor designs for applications ranging from large-surface skins to high-resolution tactile fingertip sensors. The developed theory for obtaining optimal super-resolution can guide future sensor designs.

Kurzfassung

Um Interaktionen in komplexen, sich verändernden und von Menschen beeinflussten Umgebungen wahrnehmen zu können, benötigen Roboter, analog zu biologischen Systemen, hautähnliche Wahrnehmungssensoren. Die derzeitigen hautähnlichen Sensortechnologien sind in Bezug auf Auflösung, Dynamik, Robustheit und Oberflächenabdeckung noch weit von ihren biologischen Vorbildern entfernt. Eine zentrale Herausforderung ist die Verdrahtung der Sensorelemente. In meiner Doktorarbeit untersuche ich, wie maschinelles Lernen die Entwicklung einer neuen Art von haptischen Sensoren ermöglichen kann, um solche Herausforderungen zu bewältigen. Einerseits schlage ich superauflösende, hautähnliche, taktile Sensoren vor, die die Anzahl der physischen Sensoreinheiten reduzieren und gleichzeitig eine hohe räumliche Genauigkeit erreichen. Auf der anderen Seite untersuche ich die Entwicklung von kamerabasierten haptischen Sensoren. In dieser Arbeit stelle ich vier Arten von datengetrieben haptischen Sensoren für grobe und feine Roboteranwendungen vor, die von großflächigen (Robotergliedmaßen) bis zu Sensoren mit kleinen Abmessungen (Roboterfinger) reichen. Darüber hinaus schlage ich eine Theorie der Superauflösung vor, um Sensorentwürfe auf allen Ebenen anzuleiten, vom Hardwaredesign (Material/Struktur/Transduktion) über die Datenerfassung (real/simuliert) bis hin zu den Signalverarbeitungsmethoden (analytisch/datengesteuert).

Ich untersuche zwei Entwürfe für großflächige, grob aufgelöste Sensoren, z. B. für robotische Gliedmaßen. **HapDef** bringt einige wenige Dehnungsmessstreifen auf einer großen gekrümmten Oberfläche an, um die Verformung über die gesamte Oberfläche zu messen. **ERT-DNN** umhüllt eine große Fläche mit einem Stück mehrlagigen leitfähigen Gewebes, das seine Leitfähigkeit je nach Kontakt variiert. Ich konzipiere außerdem zwei Ansätze für kleinräumige, fein auflösende Sensoren, z. B. für Roboter-Fingerspitzen. **BaroDome** bettet einige wenige Barometer in ein weiches Elastomer ein, um durch äußere Einflüsse verursachte interne Druckänderungen zu messen. **Insight** umschließt eine hochauflösende Kamera, um eine weiche Hülle von innen zu betrachten.

Generell muss ein inverses Problem gelöst werden, wenn man versucht, hochauflösende Messungen mit wenigen physikalischen Messeinheiten zu erzielen. Ich entwickle Frameworks für maschinelles Lernen, die für die Lösung dieses inversen Problems geeignet sind. Sie verarbeiten verschiedene Sensor-Rohdaten und extrahieren in der Praxis hierarchische haptische Informationen. Die Methoden des maschinellen Lernens stützen sich auf Daten, die entweder von einem automatisierten robotergestützten Stimulationsgerät erfasst oder mithilfe von Finite-Elemente-Methoden synthetisiert wurden. Ich baue mehrere physikalischen Stimulationsgeräte und Finite-Elemente-Modelle, um umfangreiche Daten zu sammeln. Ich schlage Frameworks für maschinelles Lernen vor, um Daten aus verschiede-

nen Quellen zu kombinieren, die gut genug sind, um mit dem Rauschen in realen Daten umzugehen und eine gute Verallgemeinerung von gesehenen zu ungesehenen Situationen zu ermöglichen.

Bei der Entwicklung meiner Prototypsensoren bin ich immer wieder mit Designentscheidungen konfrontiert worden. Um meine Entwicklungen zu unterstützen und künftige Forschungen zu leiten, schlage ich eine einheitliche Theorie mit dem Konzept der Taxel-Value-Isolines vor. Sie erfasst die physikalischen Effekte, die für die Superauflösung erforderlich sind, verknüpft sie mit allen Teilen des Sensordesigns und ermöglicht es uns, sie quantitativ zu bewerten. Die Theorie bietet eine Erklärung für die physikalisch erreichbaren Genauigkeiten bei der Lokalisierung und Quantifizierung von Kontakten auf der Grundlage der Unsicherheiten, die durch das Messrauschen in den Sensoreinheiten entstehen. Die theoretische Analyse zielt darauf ab, die beste Leistung vorherzusagen, bevor ein physischer Prototyp gebaut wird, und hilft bei der Bewertung des Hardwaredesigns, der Datenerfassung und der Datenverarbeitungsmethoden während der Implementierung.

Diese Arbeit präsentiert eine neue Perspektive für die Entwicklung haptischer Sensoren. Unter Verwendung von maschinellem Lernen als Ersatz für die gesamte Datenverarbeitungspipeline präsentiere ich mehrere haptische Sensordesigns für Anwendungen, die von großflächigen hautähnlichen Sensoren bis hin zu hochauflösenden taktilen Fingerspitzensensoren reichen. Die entwickelte Theorie zur Erzielung einer optimalen Superauflösung kann als Leitfaden für zukünftige Sensordesigns dienen.

“People who are really serious about software,
should make their own hardware.”

by Alan Kay

“Every privilege comes with responsibility,
every opportunity with a duty,
and every possession with a commitment.”

by Branco Weiss

“一墨一笔, 一介书生, 足矣。”

致 孙环波

Acknowledgments

First and foremost, I sincerely express my gratitude and respect to my supervisor, Dr. Georg Martius, for guiding me through this fascinating field of research, for his continuous support, insightful feedback during my Ph.D. study, and for giving me the freedom to explore my research ideas independently. I would also like to thank Dr. Katherine J. Kuchenbecker for serving on my thesis advisory committee, her long-term support, fruitful discussions in our joint projects, her constructive advising on the thesis preparation and future career plan. My sincere thanks go to Dr. Hendrik Lensch for his effort to check my Ph.D. study progress as one of my thesis advisory committee members, and for his insightful suggestions and comments on my dissertation contents.

I want to emphasize the thanks to Dr. Hyosang Lee for bringing me along on his incredible sensor development journey, for the enjoyable collaboration experience he was involved in, and for his supportive advice on my future career choices. Besides, I wish to thank Dr. Adam Spiers and Dr. Jonathan Fiene for their insightful thoughts and discussions in our collaboration project. I also wish to thank all other collaborators: Dr. Bernard Javot, Felix Grimminger, Dr. Gokhan Serhat, and Hyunkyu Park.

Special thanks to all nice colleagues at the Max Planck Institute for Intelligent Systems for creating a friendly working environment. I sincerely appreciate and greatly benefit from the time we spent together. I would also like to thank Dr. Leila Masri for her continuous support in the Ph.D. program I attended, the International Max Planck Research School for Intelligent Systems.

Moreover, I want to thank Yijie Gong, An Mo, and Dr. Yitian Shao for providing feedback on the write-up of Chapter 2 and Chapter 3, Dr. Gokhan Serhat, Hyunkyu Park, and Dr. Hyosang Lee for Chapter 4, Dr. Pavel Kolev for Chapter 5, Cansu Sancaktar, Dr. Arash Tavakoli, and Iris Andrussow for Chapter 6.

Finally, I would like to thank my family and friends (Haolong Li, An Mo, Kai Chen, Yu Zhang, Dr. Jiachen Zhang, Dr. Junjun Huang, Dr. Xiaodong Deng, Dr. Tao Chen, Dr. Xiangzhou Tan, Dr. Shuaidong Huo, Dr. Cheng Li, Dr. Long Ji, Dr. Tianxiang Huang, Yang Wang, Can Yang, Deru Song, Xiangru Li, Yawen Guan, Xinhu Liu, etc.) for their constant support throughout my overseas study period and my life in general.

The work presented in this thesis was supported in part by the China Scholarship Council (CSC), the Max Planck Society, and the International Max Planck Research School for Intelligent Systems (IMPRS-IS).

Huanbo Sun
June 9, 2022

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We are witnessing a rapid evolution of robot technologies to perform practical physical tasks in various application areas [1, 2, 3, 4]. To robustly and safely react in complex, changing, and especially human-involved environments, a robot must be able to perceive when, where, and how its body is contacting other things. Due to the visual occlusion and the small scale of deformations during the interaction, robots need touch-sensitive skin in addition to well-developed visual feedback. Comparatively, technologies for haptic feedback are still in the infancy phase and underdeveloped. Because skin-like haptic devices covering a broad range of areas are more challenging to design, with additional crucial requirements on robustness due to constant and direct external interactions. The wiring complexity increases along with high-resolution requirements.

There have been many efforts to create haptic sensors [5] that can quantify contact across the surface of a robot: successful designs produce measurements using resistive [6, 7, 8, 9], capacitive [10, 11, 12], ferroelectric [13], triboelectric [14], hall-effect-based [15], and optoresistive [16, 17] transduction approaches. Most recently, vision-based haptic sensors [18, 19, 20, 21, 22, 23] have demonstrated a new family of solutions, typically using an internal camera that views the soft contact surface from within. Most of these sensors tend to focus on an individual goal of improving precision or sensitivity, increasing taxel density, or enlarging the sensed area rather than comprehensively optimizing the **system robustness** and the usability of the sensed **haptic information**. This situation, in turn, has resulted in sensors that are either fragile, insensitive, inaccurate, bulky, or costly and hampers their widespread application and advancement in the robotic field.

Recent developments have shown that machine-learning-based approaches are especially promising for creating dexterous robots [2, 24, 25]. In such self-learning scenarios and real-world applications, the need for extensive data makes it particularly critical that sensors are robust and keep providing good readings over thousands of hours of rough interaction. On the one hand, machine-learning-driven robots put forward higher requirements for sensor robustness; on the other hand, machine learning also opens a new possibility for tackling the haptic sensing challenges in presenting unified and intuitive haptic information. Data-driven machine-learning methods can replace hand-crafted numeric calibration procedures with end-to-end mappings learned from raw sensor data and provide understandable haptic information, such as forces in newtons and contact

locations in Euclidean spaces.

In this thesis, we explore the approaches of creating robust and durable haptic sensors and develop adequate machine-learning frameworks to extract useful haptic information from raw sensor values.

1.1 Motivation

Existing robots consisting of *stiff* links like LWR (Kuka) [26], Panda (Franka Emika) [27], and APAS (Bosch) [28] are favorable products in the market due to their compact configuration, reliable accuracy, and high durability. With increased kinematic degrees of freedom (DoF), robots like Cheetah (MIT) [29], ANYmal C (ANYbotics) [30], Spot and Atlas (Boston Dynamics) [31], and Shadow Dexterous Hand (Shadow Robot) [32] inspire considerable media coverage owing to their functional dexterity and bionic appearance. They are currently in their infant phase of stepping into the mature robotic market. *Soft* robots like TentacleGripper (Festo) [33], Octobot (Harvard) [34] represent an exciting, relatively emerging area of research because of the favorable properties of soft materials for manipulating objects [35], for safer interactions in human environments [36], and to limit the instantaneous impact forces caused by unforeseen collisions in robotic systems [37]. Robots are rapidly evolving towards higher integration, higher dexterity, and more softness, where sensors also have to be adaptively designed following this evolution to facilitate the research in robotics.

A sense of vision allows **active** visual perceptions of the world from different viewing points and helps robots to navigate, avoiding collisions, breaking things, or even hurting a person. When it comes to manipulating things, such as packing a box, cleaning up a room, and helping a person, robots will need to reach out and **interactively** touch something. The touching details between objects are almost not visible, and the contacts can happen in a more distributed way over the whole body rather than centralized. A sense of touch (haptic feedback) can help process different kinds of physical stimuli to tell whether something is touching and how big an area it is covering, whether the interaction is stretching or shaking, or whether there is heat transfer, etc.

Over the last four decades, many haptic sensor schemes have been designed and are summarized in [5, 38, 39]. Since 1980s, efforts were made to create haptic feedback with high taxel density in a very focused area for end-effectors of robotic arms/manipulators [40]. Intrinsic six-dimensional force sensors were integrated into robot joints to infer contact on stiff robot links [41], which is insufficient due to unmodeled dynamic forces and inertial interference. A few early sensors were designed to cover robot arms that supported the motion planning research in unstructured environments [42, 43]. Later on, two broad categories of haptic sensor design are formed based on the sensing size and spatially resolved information: small-scale finely resolved sensors focusing on small areas and large-scale coarsely resolved sensors covering large areas [44].

Agile and dexterous robots may find themselves in situations where they have to

constantly contact external objects like manipulating objects in hand, such as fingertips or gripper end-effectors. They request more delicate haptic information like detailed texture and slippage in a focused area. Large-scale haptic sensors enable the research of currently favorable *stiff* robots in unstructured environments. With an emphasis on the need for large surface sensing for contact or collision detection during robot motion, the haptic information of location, strength, and number of contacts is more important. It coincides with the distribution of mechanoreceptors (MRs) in human skin, where 29.3% of MRs are in the finger area, 40.9% in the glabrous area of the hands, only 3.9% in the limbs (without hands), and 5.3% in the whole back and chest [45]. Delicate haptic information can be captured and delivered in many forms, such as vibration [46], deformation [18, 47], undirected pressure distribution [7], and directional force distribution [48, 49], as well as temperature [6, 50, 51]. For robotic manipulation tasks, a directional force distribution is one of the preferred forms of contact information. It describes the location and size of each contact region and the local loading in the normal and shear directions in one shot [38].

Haptic sensors transduce from physical phenomena, e.g., force-caused displacement, to an electrical signal encoding contact information. However, it is nontrivial to decode raw sensor readings to an intelligible representation of the original causes. Different approaches have been developed for mapping raw sensor outputs to real-world values, such as displacements in meters [18, 47] or contact forces in newtons [48, 49]. A linear mapping with a stiffness matrix was favorable in the past due to its simplicity [49, 52, 53]. However, this assumption is rarely valid in sensors with nonlinearity, strong hysteresis, or damping properties due to the skin materials [54]. Methods like analytical inverse compensation can partially take care of the hysteresis-caused nonlinearity problem [55]. The problems of mechanical nonlinearities and hysteresis are more noticeable in soft materials that may be negligible in their stiff counterparts [56]. Moreover, the mapping task for global deformations caused by focused loads can be tackled through neither linear mapping nor inverse compensation but finite element methods (FEMs) with detailed modeling of material properties [48]. Machine-learning methods, e.g., artificial neural networks (ANNs) [57] and Gaussian processes (GPs) [58], show their prominence in solving mapping problems that involve both nonlinearity and hysteresis, as well as non-locality. In a supervised learning manner, machine learning models can map raw sensor data to real-world calibrated values directly, without any unnecessary intermediate steps. Labeled sensor data can be collected either in a designed test environment or a robotic learning system. In addition, machine learning can help to extract more hierarchical information from raw sensor readings, such as object hardness and curvature [59].

Machine learning is a branch of data analysis methods that automates model building and enables intelligent systems to learn from data, identify patterns and make decisions with minimal human intervention. Its rise can be traced back to the 1940s, when the computational model of *Neural Networks* was introduced in 1943 [60], and the first *Hebb learning rule* for neural networks was defined in 1949 [61]. After a silent period around two decades awaiting the development of computational power, the famous and reliable *k*-

Nearest Neighbor (KNN) method for pattern classification was proposed in 1967 [62]. The *Back-propagation* method was invented to reduce the time for training neuron networks in 1974 [63]. The method of *Decision Trees* was introduced in 1986 [64]. Moreover, the originator of the widely used deep *convolutional neural network* (CNN) was invented to recognize handwritten digits in 1989 [65]. Later on, methods like *Boosting* [66], *Support Vector Machine* (SVM) [67], and *Random Forests* [68] were proposed and gradually took the shape of state-of-the-art machine learning techniques. Around 2009, a large visual database named *ImageNet* was created as the catalyst of the modern artificial intelligence boom [69]. *AlexNet* [70] was invented to push *ImageNet* classification accuracy by a significant stride compared to traditional methodologies. And *ResNet* [71] was recognized to be the first model to create very deep architectures. These modern novel networks inspire researchers to explore more efficient deep learning techniques.

With the modern advanced machine learning techniques, portable computing devices with strong computing power, and the flexibility of automatic data collection in large databases, data-driven hardware design concepts become more attractive. Because they help avoid complicated calibration and heavy hand-crafted post-processing steps and provide information in an end-to-end and interpretable way.

This thesis aims at making contributions to the development of machine-learning-driven haptic sensors for both small-scale finely resolved robotic parts and large-scale coarsely resolved robotic parts. It thus facilitates robotic researches in complex, changing, and human-involved environments. We introduce the state-of-the-art haptic sensors in Section 1.2 and the customarily used machine-learning methods for haptics in Section 1.3. Then we introduce our contributions to the haptic sensor community and the layout of this thesis in Section 1.4.

1.2 State-of-the-art Haptic Sensors

Haptic sensors evolve quickly along with the advancement of robotic techniques and the outbreak of artificial intelligence. Some review papers present thorough overviews of existing haptic sensors [5, 38, 39, 44, 56, 72]. They have different review focuses, which will be detailed below. Depending on the task, different specifications of design criteria for haptic sensors have been reported by Yang et al. [5] for biomedical engineering, by Francomano et al. [73] for prosthetic hands, by Dahiya et al. [74] for large haptic systems, by Dahiya et al. [44] for humanoid robots, and by Balasubramanian et al. [75] for dexterous manipulators, as well as by Chin et al. [56] for soft robots. Regarding the design principle, the transduction from physical phenomena to an electrical signal can be constructed differently, like measuring the change of resistance [76, 77], capacitance [78, 79], electrical charge [80, 81], and optical characteristics [18, 22]. We discuss common design criteria for haptic sensor design in Section 1.2.1 and describe the basic transduction methods in Section 1.2.2. In addition, we introduce several representative haptic sensor designs in Section 1.2.3.

1.2.1 Design Criteria

I will first introduce performance measures and criteria that can guide the development and specification of sensors in the following:

Spatial resolution and complexity of wiring The spatial resolution describes the number of taxels integrated into a tactile sensing array. The high spatial resolution requires numerous sensing taxels, which unavoidably leads to complex wire connections, long acquisition and processing time, and high sensitivity to external electromagnetic noises and crosstalk. Referring to human skins [45], a hierarchical arrangement of spatial resolution for different applications is favorable, such as the lower spatial resolution of about 5 mm for large surfaces (limbs) and finer around 1 mm for end-effectors (fingertips) [44]. High spatial resolution is often required in contact pattern recognition and delicate manipulation scenarios. The complexity of wiring can be reduced by methods like multiplexing and serial communication protocol.

Sensitivity and dynamic range The sensitivity of sensors defines the smallest detectable external impact, and the dynamic range frames the sensors' measurement range from the smallest to the biggest. Usually, the dynamic range shrinks when the sensitivity increases due to the materials and structures of currently available sensors. Suggested in [44, 72], a force sensitivity range of 1-1000 gram and a dynamic range of 1000:1 are desirable in robotic applications. Combining multiple haptic sensors with different sensitivities and dynamic ranges can help cover a larger measurement range and ensure high sensitivity. High sensitivity is usually desired in light touch detection and fragile object manipulation tasks.

Signal frequency There is an essential difference between signal frequency and frequency response that are often mix-used. Signal frequency describes how fast the haptic sensor responds to external contact and delivers such signal to a central processing unit. However, frequency response is a signal format, vibration, measured by the sensors, and different sensors may have different frequency responses. High signal frequency offers near real-time haptic feedback but is not always preferred as it burdens energy and computational resources. Different frequency responses of sensors can help to discriminate quasi-static and dynamic contacts, such as quasi-static force distribution measurement (low vibration frequency) [76] and dynamic (incipient) slippage detection (high vibration frequency) [82]. As a reference for frequency responses, human MRs can vibrate and respond to frequencies up to 400 Hz [83].

Elastic hysteresis Elastic hysteresis is a material effect in that energy is dissipated as internal friction (heat) during contact loading and unloading, and the shape recovery after unloading has a time hysteresis. In sensor design, the elastic hysteresis should be as low

as possible, as it damps frequency response and has long-time memory of the previous impact. In contrast, favorable soft materials have more substantial hysteresis than stiff materials. Besides, methods like inverse compensation [55] and machine-learning [57] can relieve this effect.

Mechanical properties A haptic sensor should be ideally flexible to be attached to any shaped surfaces or designed in any 3D shape. Moreover, a sensor should be durable for highly repetitive usage without a performance drop. The sensor surface is supposed to have mechanical compliance with external contacts. The surface friction coefficient should be adapted to applications: a high coefficient is needed in manipulation devices to hold the object in hand [77], and a low coefficient is desired in exploration procedures [84]. The structure with papillary ridges like fingerprints can also increase the friction coefficient, apart from the material itself.

Data transmission and interpretation The electrical signals transduced from haptic sensors flow into the central processing unit with possible intermediate steps of analog-to-digital converting, serial paralleling, data preprocessing (data selection, local computation), and haptic information interpretation. The host robot should allow attention-triggered sensor data selection to save processing resources and prioritize them with different transfer and processing rates. The haptic information can ideally be locally preprocessed and interpreted in the needed format, rather than flushing unnecessary raw data into the host processing unit.

1.2.2 Transduction Methods

Haptic sensors make use of the physical/mechanical nature to transduce contact information to digital signals. Transduction methods vary in a wide range, and we will introduce some commonly used ones in the following:

Piezoresistive types Piezoresistive sensors transduce mechanical deformations in the form of electrical resistance changes. Several classical technologies use the material's piezoresistive property to design haptic sensors: direct use of discrete force sensing resistors (FSR) to quantify contact location and strength [85, 86]; a mixture of electrically conductive carbon particles in non-conductive elastomer to construct a soft and conductive rubber [87]; moreover, a hierarchical design of multilayer-structured electrically conductive fabric to distribute variable resistance spatially [88, 89]. Piezoresistive sensors are validated to be widely accessible and low cost with simple electronics and manufacturing processes, incredibly widely adopted in micro-electro-mechanical systems (MEMS) and silicon-based designs [90, 91]. They are usually in a flexible mechanical structure, have good sensitivity, and offer low noise signals. Conversely, piezoresistive sensors are often

non-linear and respond both to temperature and moisture. Apart from the hysteresis effect, they have high power consumption and tend to be fragile to shear force.

Capacitive types Capacitive sensors measure capacitance changes between two conductive plates separated by compressible dielectric materials when external contacts are exerted. Capacitive technologies are popularly used in our daily lives, such as touch screens, and have been well investigated. Based on MEMS and Si micromachining, capacitive sensors can be made very compact and small, typically in constructing dense sensor arrays. Many customized capacitive sensors are available, such as the haptic skin for the Robotiq robot gripper [79], the iCub humanoid robot [92], and the PR robot grippers [93]. Moreover, commercialized products are maturely available such as DigiTacts from Pressure Profile Systems [94] and AD7147 from Analog Devices [95]. Capacitive sensors are robust and sensitive, can be flexibly manufactured, and have higher bandwidth than piezoresistive sensors. Similar to the drawbacks of sensitivity to temperature change, non-linear response, and hysteresis in piezoresistive sensors, capacitive ones may bring electromagnetic noise and stray capacity problems with themselves.

Piezoelectric types Piezoelectric-effect-based sensors generate electrical charges when mechanical energy is captured. Among many piezoelectric materials, quartz crystals and ceramics show better piezoelectric properties, while polymers like polyvinylidene fluoride (PVDF) are preferred with prominence in flexibility and chemical stability [96]. Piezoelectric sensors based on PVDF were reported to have a frequency response range of 1 Hz to 7 kHz [80, 81]. These sensors have a faster dynamic response with very high bandwidth for dynamic loading detection than capacitive sensors. However, they can not detect static loading and also have temperature dependence.

Quantum tunnel composite (QTC) types QTC sensors make use of the quantum tunneling effect to change the material property from insulator to conductor upon external loading is exerted. With a simple structure design, the QTC-based sensors can discriminate normal and shear forces with high sensitivity and high dynamic range [97]. They are very energy-efficient, as they consume power only when contacts exist [98, 99]. Due to the manufacturing complexity and high cost, QTC materials are not widely accessible. Only Peratech [100] supplies commercial products.

Optic types Optical sensors measure light intensity change caused by mechanical deformation. Commonly used optic sensors have four essential elements: light emitter, light detector, deformable light-transmission-medium, and support structure. There are several well-developed techniques for detecting the light intensity change. Photodetectors measure the residual intensity of light that transmits from infrared light-emitting diodes (LED) and through the light-transmission-medium [101]. Fiber Bragg grating constructs optical fiber to vary the fiber core's refractive index in correlation with the strain change

and measures the reflected light wavelength to reveal that change due to either external loading or temperature change [102]. Moreover, vision-based haptic sensors typically use an internal camera that views the soft contact surface from within [18, 49]. Optical sensors are favorable due to their high sensitivity and immunity to electromagnetic interference [103]. They can gain very high spatial resolution with extremely densely placed receptors, e.g., high-resolution cameras, at the cost of speed, power consumption, and computational costs. Tactile sensors need to be easy to produce from inexpensive components to facilitate their widespread adoption. Imaging components are remarkably cheap these days due to the standardized Complementary-Metal-Oxide-Semiconductor (CMOS) manufacturing procedure, making vision-based sensors competitive. Moreover, nowadays, computation is fast enough to perform image processing, in classical ways, with machine-learning methods.

1.2.3 Representative Haptic Sensor Variants

Many review papers present thorough overviews of existing haptic sensors for various applications [5, 38, 39, 44, 56, 72]. Here we introduce several representative ones with two categories: small-scale but often finely resolved sensing and large-scale coarsely resolved sensing.

Small-scale Finely Resolved Sensing TakkTile [104] molds a compact barometer inside silicon rubber to form a tactile sensing cell. A focused area can be equipped with finely resolved sensing ability by integrating multiple such cells, such as TakkStrip [105] and iHY hand [106]. A sensor based on the same physical effect is also studied in Chapter 5 Kõiva et al. [77] incorporate a non-flexible pattern of the electrodes on a 3D curved layer and piezoresistive rubber on a second layer to design a simple setup for 3D surface sensing and a soft compliant sensing surface. Syntouch BioTac [6] is a multimodular sensor for high-resolution tactile sensing. It uses 19 electrodes encased in a conductive liquid-filled elastic skin for tactile sensing on a human-finger-sized device. It integrates different tactile sensing modalities, such as thermal, pressure, and vibration sensing modules. However, the BioTac is costly (15 000 USD) and is said to be fragile and have sensing repeatability issues.

GelSight [18] is one of the earliest vision-based sensors, which has a thin reflective coating on top of a transparent elastomer layer supported by a flat acrylic plate. Perpendicular lighting allows tiny surface deformations to be detected using photometric stereo techniques. Further developments of this approach increased its robustness (GelSlim [19]), achieved curved sensing surfaces with one camera (GelTip [107]) and with five cameras (OmniTact [108]), and included markers to obtain shear force information [109]. A different technique based on tracking tiny beads inside a transparent elastomer is used by GelForce [49] and the Sferrazza and D'Andrea sensor [20] to estimate normal and shear force maps. ChromaForce uses subtractive color mixing to extract similar data from

deformable optical markers in a transparent layer [110]. The TacTip [22] sensor family uses a hollow structure with a soft shell, and it detects deformations on the inside of that shell by visually tracking markers.

Large-scale Coarsely Resolved Sensing HEX-o-SKIN [11] integrates a proximity sensor, an accelerometer, three normal force sensors, and a temperature sensor on a 15 $\times$ 15 mm hexagonal printed circuit. It allows covering a surface, e.g., of a robot exoskeleton, with multiple HEX-o-SKIN chips forming a dense array. In this way, a large surface can be covered. However, the robustness of the system might be challenging.

TacCylinder [22] is a camera-based system. It is shaped like a cylinder with a tube through its center, which holds a camera and a bulky catadioptric mirror system to capture the motion of markers painted on the internal flexible shell surface. The sensor has a dimension of 63 $\times$ 63 $\times$ 82 mm and delivers comprehensive information about the deformation of the soft cylindrical surface. The surface shape is restricted, and a new shape requires an adaptation of the optical system. Additionally, the inside of the mechanical part needs to be empty/visually transparent for this method to be applicable. Similar works have also been implemented in [111, 112].

Lee [76] uses stretchable conductive materials (skin) with a few electrodes assembled on the skin boundary and measures all combinations of pairwise conductivity. The contact information of deformation is determined by anisotropic electrical impedance tomography (EIT). Only 16 electrodes are required on the skin boundary with a skin size of 40 $\times$ 100 mm. However, high computational costs arise, requiring special hardware. Later on, EIT-oriented sensors have been further developed by using different materials, such as conductive fabric [88].

1.3 Machine Learning in Haptic Sensors

Haptic sensors capture contact information based on various physical phenomena, such as changes of resistance [7, 8, 77]. The inverse problem of inferring haptic information in formats of real-world values from the raw sensor values is nontrivial to solve, e.g., resistance changes to forces in newtons [48, 49]. There are models, in many cases, and analytical calculations that can be done. However, these models are inaccurate because of nonlinearity, inhomogeneity, and damping properties [54]. These properties are more prevalent in soft materials that have advantages in tactile sensor design. Various formats describe the haptic information in real-world values, such as contact location, strength, area, direction, motion, posture, temperature, etc. The inverse mapping functions are typically unknown and nontrivial to represent analytically. Data-driven methods such as machine learning show promising results in approximating these functions. We discuss the most common real-world values for describing haptic information in Section 1.3.1 and describe the basic machine-learning methods successfully employed in haptic sensor designs in Section 1.3.2.

1.3.1 Haptic Information

When thinking about which haptic information a sensor can provide, it is instructive to discuss different kinds of contact events. One can distinguish the approaching, establishing, holding, in-position moving, and leaving phases. Approaching is the state before the contact establishes. Proximity sensors [42, 113] are typically used to provide the relative geometrical relation to an object. Imagine an object is grasped by a robotic hand. When physical contact occurs or ceases to exist, haptic sensor values change abruptly. Depending on the material, high-frequency vibrations can occur [46]. During a firm grasp, i.e., a quasi-static holding, we can extract several properties about the interaction: center of contact, contact area/shape, and normal/shear contact strength. A different situation is when the hand is moving relative to the object's surface. It requires sensors to provide the information mentioned above in real-time continuously. All the information can be combined to extract a higher level of haptic information, such as the object's overall shape, texture, roughness, hardness, material properties, etc.

1.3.2 Machine-Learning Methods

Machine-learning methods show high adaptability in solving nonlinear problems and extracting useful information from various data types [62, 65, 67, 68, 114, 115, 116, 117]. The basic rule shared between different machine-learning methods is that similar data share similar properties. There are three major branches of machine-learning methods: supervised learning, unsupervised learning, and reinforcement learning. In this thesis, I only consider supervised learning:

$$argmin_{f \in \mathcal{F}} E_{x,y \sim D} L(f(x), y). \tag{1.1}$$

We search for a function f in the space of F that yields minimal expected loss L under data (x, y) from the data distribution of D. In practice, we can typically not find the best function f^* because we cannot search through all functions in F. Instead, we aim to find a good enough solution $\tilde{f}$. For haptic sensors, we need to find a $\tilde{f}$ for solving the inverse problem that maps raw sensor values (x) to intuitive representations (y). Instead of writing $\tilde{f}$ by hand, we can use machine-learning methods, as indicated by Eq. 1.1. The research field on this topic is large, and many function spaces and different optimization methods have been developed. In my thesis, I focus on implementing regression tasks, i.e., where y are real values/vectors in $\mathcal{R}^n$. Another typical case would be the classification task with y being discrete.

A regular procedure for machine-learning methods used in haptic sensors is to collect copious data, including raw sensor values (x) and labels (y). Labels can be generated by well-known quantification standards such as force in newtons [57] or prior human knowledge to describe the object property such as smoothness [118]. The data is then split into training-validation-test subsets. The training dataset is used to train a machine

learning model with variable structures and tune-able hyperparameters. The validation dataset is used to validate whether the trained model has overfitting problems, which may suggest a reconfiguration for the trained model. The test dataset is then used to evaluate the sensor's overall performance and the machine-learning method. It reveals the performance of the sensor in running time.

In the following, I introduce several representative machine-learning methods that are favorably used in haptic sensors.

k-nearest Neighbors Algorithms (kNN) kNN [62] is a non-parametric method that compares the incoming test data samples with k nearest training samples and outputs an averaged or weighted label. It can be used for either classification such as object smoothness levels [119] or regression such as contact locations and force strengths [120].

Support Vector Machine (SVM) SVM relies on non-linear kernel functions to map the input (raw sensor values) into a high-dimensional feature space and find a hyperplane defined by support vectors with maximum margin, such that all data points are still correctly classified [121]. A similar procedure can be done for regression problems, called support vector regression [120].

Random Forests (RF) RF builds numerous decision trees at the training time and returns a class selected by most trees for classification or an averaged prediction of the individual trees for regression. As done in [122] for haptic applications, it can generate reasonable predictions while requiring little structure configuration.

Gaussian Process (GP) GP is a non-parametric process with a kernel function measuring similarity between data samples. GP compares the test samples with training samples and outputs a mean estimate and the uncertainty represented by a one-dimensional Gaussian distribution [58].

Deep Neural Networks (DNN) DNN is a category of artificial neural networks (ANNs) that works for both regression and classification tasks [15, 6]. It consists of multiple layers of perceptrons with threshold functions between the input (x) and output (y) layers. A simple function

$$f = W_2\sigma(W_1x + b_1) + b_2 \tag{1.2}$$

performs both linear transformations (W_1x+b_1) and nonlinear transformations ($\sigma(\cdot)$). The parameters $\theta = (W_1, W_2, ..., b_1, b_2, ...)$ can be adapted using stochastic gradient descent by minimizing $E_{x,y\sim D_{train}}L(f(x),y)$. A typical loss L for regression task is the square loss $L(\hat{y},y) = \sum(\hat{y}-y)^2$. If W is a full matrix, we call them fully connected neural network (FCN) or multilayer perceptron (MLP). It was found that it is useful to encode knowledge about the structure of the problem into the structure of neural networks

and bring inductive bias to the machine-learning methods. For example, convolutional neural networks (CNNs) are the regularized versions of MLP. They are primarily used to process image-like data. CNNs utilize the hierarchical pattern in image-like data and assemble patterns using smaller and simpler patch kernels/filters. In the end, hierarchical information can be extracted at lower complexity compared to MLP [18, 20, 94, 108].

1.4 Thesis Contributions and Outline

1.4.1 Contribution

Haptic sensors are indispensable to enable robots to perceive when, where, and how their bodies are contacting other things. A common theme for haptic sensors is integrating many small sensing elements forming a grid along a flat or curved surface. Each sensing element, named taxel, is responsible for sensing interactions near its location. For typical applications, a resolution is desirable that would imply numerous taxels. It is valid for small surface sensing, e.g., at the fingertips, and for large surfaces, e.g., around limbs, as shown in Fig. 1.1. A high density of smaller taxels and thinner wires are favorable regarding high-resolution, but less ideal due to mechanical fragility, electromagnetic noise, and crosstalk between taxels. Technical challenges arise concerning the physical size of the taxels, growing manufacturing, and wiring costs. These closely relate to robustness issues in hardware design. High-fidelity haptic sensors are still not widely commercially available, and robustness issues often constrain their application.

Efforts are made to acquire as much tactile information as possible from only a few physical taxels [6, 124, 15, 125, 126]. The reduction of the number of taxels is generally possible because a taxel can monitor a tiny area and an extended patch on the surface, and multiple taxels can jointly provide information about tactile information at a certain position. The particular material properties lead to a characteristic spread of contact information to the sensing taxels. Different physical effects can be used, such as electrical resistance [6, 124], magnetic flux [15], thermo- and fluid-dynamics [126], geometric and mechanical properties [127], etc. The central idea is to solve the inverse problem of inferring haptic information from a few sensors, effectively creating high-resolution virtual taxels. It is also referred to as super-resolution sensing [128].

Haptic sensors transduce contact events into raw digital sensor values, which are typically non-linear and have no direct indication of real-world values, such as force in newtons, pressure in pascals, locations in 3D Euclidean spaces. Carefully calibrated models using numerical methods can infer haptic information from sensor readings with high-speed efficiency. However, they sometimes lack precision due to modeling assumptions and unintentional ignorance of partial real-world aspects. Data-driven methods like machine learning are particularly suited to such systems because they learn a suitable mapping function from data, which can be very complicated and contain all real-world aspects of the physical sensor.

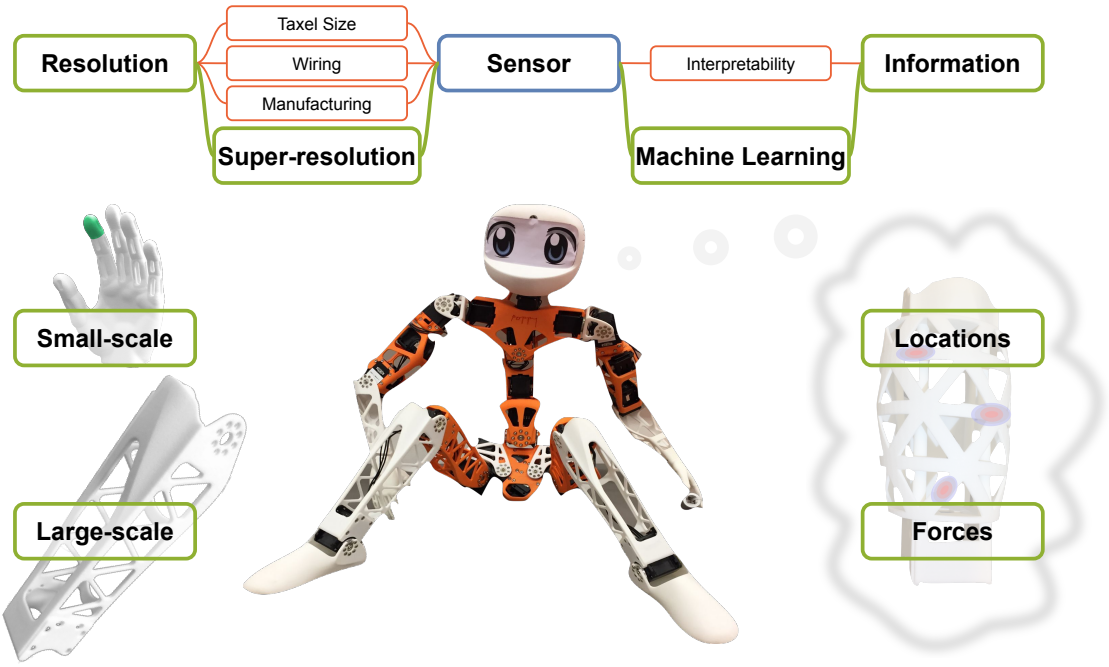

Figure 1.1: Challenges in haptic sensor design for small and large surfaces: robots like Poppy [123] need haptic feedback for small-scale (green highlighted fingertip — middle left) and large-scale (white limb — bottom left) surfaces. This thesis focuses on creating robust haptic sensors with super-resolution property that reduces taxel number, wiring, and manufacturing costs. Machine-learning methods are employed to interpret non-linear raw sensor values into real-world values and timely deliver the host robot with haptic information, such as contact locations and forces.

Thus, this thesis aims to tackle two critical tasks. The first is to create robust haptic sensors for small-scale and large-scale surfaces with a few sensing elements and wires. The second is to infer real-world haptic information from raw sensor readings in an end-to-end data-driven manner. By jointly considering requirements on spatial accuracy, force precision, unconstrained geometry, complexity in manufacturing, durability for long-term use, and usability of haptic information, we developed two sensors for large-scale coarsely resolved applications and two sensors for small-scale finely resolved applications. These four sensors are data-driven and use machine learning to infer contact information such as contact location and force strength.

There are two strategies tested for large-scale sensing. One is named **HapDef**, and it sparsely attaches a few sensor elements (strain gauges) on a large curved surface to measure the deformation and uses machine learning to infer contact information over the whole surface, as shown in Fig. 1.2 red boxes. Another is called **ERT-DNN**, and it wraps a large surface with a whole piece of conductive fabric, which varies its conductivity depending on contact force, as shown in Fig. 1.2 green boxes. There are only a few

electrodes connected to the fabric to read the spatial conductivity distribution. A machine-learning method is used to interpret real-world haptic values from the raw measurements. For small-scale sensing, there are also two approaches proposed. One is called **BaroDome**. It sparsely embeds a few sensor elements (barometers) inside a soft elastomer to measure internal pressure changes and uses a machine-learning method to predict contact details, as shown in Fig. 1.2 purple boxes. Another is named **Insight**. It encloses a high-resolution camera to view a soft shell from within and uses a machine-learning method to infer force distributions, as shown in Fig. 1.2 brown boxes. Moreover, I proposed a unified **theory** for explaining the super-resolution working mechanisms and giving guidance on sensor designs at all levels, ranging from hardware design (material/structure/transduction), data collection (real/simulated), and signal processing methods (analytical/data-driven). It bridges sensor designs (hardware) with machine learning (data-processing) in a closed-loop, as referred in Fig. 1.2 blue circle.

1.4.2 Outline

Machine-learning-driven haptic sensors require a data collection procedure. The data can be either acquired by a designed testbed, collected in a fully integrated robotic system, or synthesized using finite element methods. Data from different sensors have various properties. Thus, we need to adjust machine learning structures to them respectively and solve the tasks of inverse mapping from raw sensor values to real-world values adequately. We have different focuses for the four sensor types introduced in this thesis due to the individual data property. The thesis will introduce each sensor with its problems and the solutions we proposed. It is organized as follows. We will first introduce the **HapDef** project in Chapter 2 and Chapter 3. The introduction of the **ERT-DNN** project is stated in Chapter 4. Chapter 5 introduces the **Theory** project followed by the contents of the **BaroDome** project. Chapter 6 introduces the **Insight** project. We close with a discussion in Chapter 7. We include a general introduction about the scientific contributions of each chapter and clarify the respective author contributions in the following.

HapDef opts for a few small-sized physical sensor elements (strain gauges) measuring internal deformations of stiff and shell-shaped links that form a sparse sensor configuration with super-resolution property. Super-resolution describes a sensory capability that transcends the sampling limit set by discrete receiving elements (a few small attachable strain gauges) [127]. It offers a couple of conceptual advantages: the system is robust against environmental impacts because the sensor elements can be placed inside the structure; the surface shape can be freely designed; only a few channels need to be read out, reducing the energy consumption and increases the data rate. In this pipeline, we make the following contributions:

- We propose a new way of implementing a large-surface haptic system by placing a few strain gauges to measure internal deformations of a robotic limb and infer the single-contact information thereof. On the theory side, we provide a method

Figure 1.2: Thesis contribution overview. Red boxes describe the project **HapDef**. When two fingers contact the 3D curved skeleton, the deformation is inferred from ten sensor elements (red dots), shown as an unfolded 2D plan in terms of a color map. The contact locations and forces are then extracted accordingly. Red crosses and yellow pluses are ground truths and predictions, respectively, with variant sizes indicating different forces. Green boxes describe the project **ERT-DNN**. A few flexible and conductive sewing wires (green dots) are sewed on the conductive fabric. There is a conductivity change upon object contacting, as shown in the orange shade around the contact area. A machine learning model is built to infer the contact locations and forces for multiple contacts from the raw measurements. Purple boxes show the project **BaroDome**. A few barometers are sparsely embedded in a full soft elastomer to measure internal pressure changes. A machine learning pipeline is established to predict contact information. Brown boxes describe the project of **Insight**. A high-resolution camera is enclosed to view a soft shell from within, and machine learning is used to infer the force distributions over the outer sensing surface. The blue circle belongs to the project **Theory**. The theory explains the super-resolution working mechanisms. The capital Ω is the super-resolution factor.

for determining the optimal position of sensor placement based on finite element simulations. On the application side, we provide an assembly method for attaching strain gauges to a complex 3D structure surface and design a testbed with four DoFs to systematically collect single-contact data for training a machine-learning model. The content is presented in [129], and details are introduced in Chapter 2.

Author contribution: Huanbo Sun and Georg Martius conceived the method and the experiments; Huanbo Sun designed and constructed the hardware, conducted experiments, and analyzed the data; Georg Martius supervised the data analysis.

- We propose a real2sim framework to relieve the data insufficiency problem in inferring information for multiple-contact scenarios using the same hardware setting mentioned above. Only the real single-contact data can be collected. We first build a finite element simulation model to describe the mechanical behavior of multiple contacts and reformulate the model as an end-to-end machine-learning model. Then we train a model to calibrate the simulation model with the real sensor system using the trivially collectible real single-contact data, which is known as the transfer-learning procedure [130]; As a result, this haptic sensor uses raw strain gauge values and infers multiple contact information. The content is presented in [9], and details are introduced in Chapter 3.

Author contribution: Huanbo Sun and Georg Martius conceived the method and the experiments; Huanbo Sun designed and constructed the hardware, conducted the experiments, and analyzed the data; Georg Martius supervised the data analysis.

ERT-DNN covers the robot links with stretchable conductive fabric monitoring the resistance change upon the contacts exerted on, measured through a few electrodes sewed on the fabric based on the electrical resistivity tomography (ERT) technique. ERT is a method to reconstruct the resistance distribution over the conductive substrate from electrical potentials combinatorially measured between a few electrodes assembled on its boundary [124]. This design is beneficial to whole-body sensing for several reasons: the fabric is flexible and can seamlessly cover complex geometries; it is soft and can dampen external impacts for safer interactions; it can be manufactured at a low cost and is durable for long-term use. In this pipeline, we make the following contributions:

- We first validate the generalization property of our proposed real2sim framework in such a large-scale ERT-based conductive fabric sensor. The work is presented in [7], and Huanbo Sun contributes to this work with the idea of the real2sim framework. We then extend the testbed (single contact capability) complexity with the ability to automatically collect data of six contacts on an enlarged fabric sensing surface simultaneously and have a more thorough statistical evaluation on the performance of the real2sim framework taking the direct end-to-end mapping performance as a baseline. We first optimize and enlarge the sensor design and construct a new testbed with six indenters with force sensors to collect ground-truth multiple-contact data. We then build the real2sim machine-learning model (multi-physics model + transfer-learning calibration) and the direct end-to-end model separately. In the end, we statistically evaluate the performance of the real2sim framework. The content is presented in [131], and details are introduced in Chapter 4.

Author contribution: Hyosang Lee, Huanbo Sun, Hyunkyu Park, and Gokhan Serhat equally conceived the method and experiments; Hyosang Lee, Huanbo Sun, and

Bernard Javot designed and constructed the sensor and the testbed; Hyosang Lee conducted the experiments, collected and analyzed the raw data; Hyosang Lee and Gokhan Serhat built the multi-physics model; Hyunkyu Park collected the simulation data and contributed to analytical results; Huanbo Sun built the end-to-end model and the real2sim machine-learning model and quantitatively evaluated the performance under supervision by Georg Martius; Katherine J. Kuchenbecker supervised the whole project.

Theory introduces the concept of the taxel value isolines to describe sensor behavior. It describes the force sensitivity and spatial accuracies of contact position and force magnitude in the form of variances. Machine-learning methods with least-squares loss are used to minimize the variances and infer the contact information. **BaroDome** is an application of the proposed theory. It is designed with a 3D conical fingertip-sized sensor for a manipulator end effector. It sparsely embeds a few barometers inside a full soft elastomer to measure the internal pressure changes and uses machine learning to infer the contact information. In this pipeline, we make the following contributions:

- We introduce a theory for super-resolution-oriented haptic sensor design concepts to derive the distribution of accuracy (contact location and force strength) and sensitivity over the sensing surface from the sensor isolines. We evaluate the influence of different factors, such as the elastic properties of the materials, using finite element simulations. We compare three representative real sensor element types, empirically determine their isolines, and validate the theory in two custom-built sensors with barometric elements for 1D and 2D measurement surfaces. Our sensors obtain an unparalleled average super-resolution factor of over 100 and 1 200, respectively, applying machine-learning methods. Based on theoretical derivations and experimental results, we provide guidelines for future sensors design from different perspectives of material properties and structure design choices and baselines for evaluating the performance of trained machine-learning models. The content is presented in [132], and details are introduced in Chapter 5.

  Author contribution: Huanbo Sun and Georg Martius conceived the method and experiments; Huanbo Sun derived the theory, designed and constructed the hardware, developed fabrication methods, designed and conducted experiments, collected and analyzed the data; Georg Martius supervised the data analysis.

- We apply the theory mentioned above in a real small-scaled application by designing a 3D conical thumb-sized sensor for a manipulator end effector. We mold several barometers inside an elastomer and wrap the elastomer over a dome-shaped central core. This design gives a flexible outer surface to protect the sensors while also creates localized pressure measurements. We use machine learning techniques to

interprets the raw barometric values into localized resultant directional force vectors. Details about this work are in a subsection of Chapter 5.

Author contribution: Huanbo Sun, Hyosang Lee, Adam Spiers, Jonathan Fiene, and Georg Martius conceived the method; Huanbo Sun designed and constructed the hardware, developed fabrication methods, designed and conducted experiments, collected and analyzed the data; Georg Martius supervised the data analysis.

Insight belongs to a new family of solutions towards small-scale finely resolved sensing, typically using an internal camera that views the soft contact surface from within. Imaging components are remarkably cheap these days, making vision-based sensors competitive. By considering the goals and constraints from a fresh perspective, we have invented a vision-based sensor suitable for dexterous robotic manipulation. In this design, we make the following contributions:

- We present a high-resolution, soft, thumb-shaped sensor with all-around force-sensing capabilities enabled by vision and machine learning; it is durable, sensitive, accurate, affordable, and compact. The skin is made of a soft elastomer over-molded [133] on a stiff hollow skeleton to maintain the sensor's shape and allow high interaction forces without damage. It utilizes structured light and shading effects to monitor the 3D deformation of the sensing surface with a single camera from the inside. The sensor's output is computed by a data-driven machine learning approach, which directly infers distributed contact force information from raw camera readings. This design shows promise in high sensitivity, high spatial resolution, robustness, and a compliant soft sensing surface. The content is presented in [134], and details are introduced in Chapter 6.

Author contribution: Huanbo Sun, Katherine J. Kuchenbecker, and Georg Martius conceived the method and experiments; Huanbo Sun designed and constructed the hardware, developed fabrication methods, designed and conducted experiments, collected and analyzed the data. Georg Martius and Katherine J. Kuchenbecker supervised the data analysis.

In summary, this thesis contributes to several haptic sensor design strategies with different emphases for various robotic applications, ranging from large-scale stiff robot limbs to small-scale soft-stiff-hybrid robotic manipulators. I explore and propose different sensor designs by exploiting the possibilities of machine learning for data processing. The thesis also contributes to formalizing the theory of geometric super-resolution as guidelines for future machine-learning-driven haptic sensor design. We will discuss the pros/cons of these sensor designs and possible future directions for haptic sensor designs in Chapter 7.

Chapter 2

Large-scale Sensing with Sparse Configuration

This chapter introduces a novel pipeline to infer haptic information of single contact on the surface of a complex 3D structure from internal deformations measured by a few low-cost strain gauges instead of dense sensor arrays. We tackle two critical problems here: we first optimize the number and the placements of sensors based on finite element simulations, and then use machine learning techniques to infer single-contact haptic information. We validate the framework performance on an in-house modified limb of a humanoid robot [135]. We obtain 8 mm localization accuracy of single contact over the sensing surface area of 200 mm $\times$ 120 mm using only nine strain gauges (8 mm $\times$ 5 mm each).

2.1 Introduction

We are witnessing a rapid evolution of robot technologies ranging from robots with stiff links [26, 27, 28, 29, 30, 31, 32] to soft robots [33, 34]. Even though actuators and sensors have become increasingly powerful and compact, robots are still far from matching human capabilities, especially when it comes to touch sensation. Robots need such a capability to reliably learn interaction patterns for mastering real-world challenges in a long-term period [24]. For this, haptic sensors have to be robust in order to sustain long-lasting experiments. Moreover, another essential aspect of robotic hardware is its price. A low cost makes robotic technologies widely accessible and thus facilitates research.

In this chapter, we aim to provide a solution of an affordable haptic sensing system design for large-scale sensing. Related works about large-scale sensing have been proposed. In [11], multiple Hex-o-Skins chips are placed next to each other to cover the sensing surface. Conductive fabrics with stitched wires [7] or conductive rubber incorporating non-flexible electrodes [77] wrap the whole sensing surface. Cameras are placed inside the skeleton that views the soft contact surface from within [111, 22, 112]. Wiring problems and bulky physical settings make these approaches nontrivial to be directly used in real applications.

Figure 2.1: Overview of the method: Inferring single contact haptic information from a few sensors. We modify the shin part of a humanoid robot to validate our proposed idea.

This chapter provides an affordable, robust, and sufficiently accurate method for inferring haptic information on a large 3D sensing surface (a robot limb). We use a few physical sensors to measure internal deformations, rather than rely on a dense array of sensors on the robot's surface. It brings a couple of conceptual advantages. The system is robust against environmental impacts because the sensing elements are placed inside the structure. In addition to the flexibility in designing the surface shape, only a few channels have to be read out, reducing energy consumption and increasing data transfer rates. Fly in the ointment, the proposed solution depends on multiple taxels' joint response about contact at a certain position, while an individual taxel doesn't directly inform the stimulation quantity. An inference mechanism is needed to estimate the haptic information. We thus propose using machine learning methods to perform the inference efficiently. To acquire as few sensing elements as possible, we employ several optimization schemes to determine optimal sensor placement.

The contributions of this pipeline are as follows. From the theory perspective, we propose a new way of implementing a whole surface haptic sensor and provide a method for determining sensors' optimal number and position using finite element simulations. From the application perspective, we provide an assembly method for attaching the strain gauges, design a hardware system to collect data systematically, and demonstrate the sensing system on a robotic limb. The chapter is structured as follows. Section 2.2 presents the method by first giving an overview and then investigating optimal sensor placement. In Section 2.3, we present the results on the robot limb. We close with a discussion in Section 2.4.

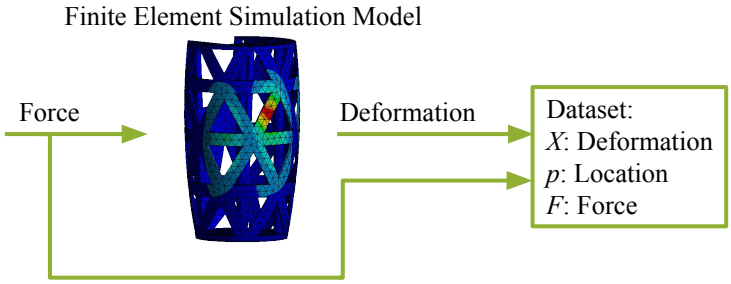

Figure 2.2: We build a finite element model to analyze the mechanical deformation property of the structure.

2.2 Methods

In this study, we reconfigure a limb of the Poppy robot with inner rigid supports and a 3D flexible outer shell. We propose to implement haptic force sensing for the flexible shell by attaching a few deformation sensors (strain gauges) on the inner side and use machine learning methods to infer single contact haptic information, as illustrated in Fig. 2.1.

We use finite element simulation (ANSYS [136]) to access data indicating how forces applied to the structure propagate into deformations measurable by the sensing elements. Based on the dataset, we answer the question of where to optimally place the sensor elements given the limited sensor number and the accuracy requirement. Given the simulated deformation patterns for different force impacts (dataset), the problem can be stated: Let $X \in \mathbb{R}^{M\times N}$ denote the deformation of N measurement points on the shell inner surface and $p \in \mathbb{R}^{M\times 3}$ the applied force's position for M different contact locations. We look for a subset of measurement locations A among all possible points N such that the contact locations can be well-inferred i. e.:

$$\arg\min_{A\subset N} \left(\mathbb{E}[\|f(X_{\cdot,A}) - p\|_2^2] \leq \delta\right), \tag{2.1}$$

where $f(\cdot)$ is a learned mapping function, δ is the tolerated error, and $\mathbb{E}$ is the expectation operator.

Our approach to approximate Eq. 2.1 comprises the following steps:

- collect a dataset of deformations based on the simulation (Section 2.2.1);
- filter possible sensor locations according to physical constraints (Section 2.2.2);
- learn a regression model to infer the contact positions (Section 2.2.3);
- decide the number and position of sensors needed for a pre-determined accuracy (Section 2.2.4).

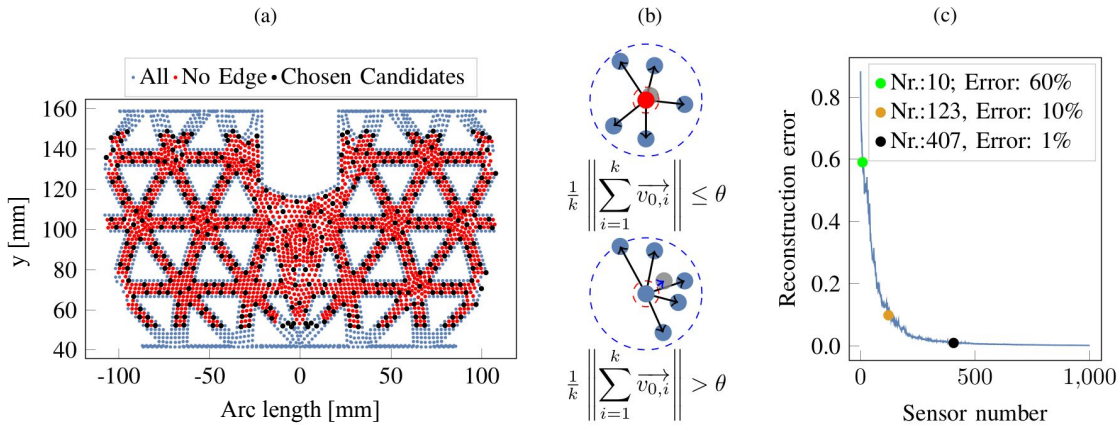

Figure 2.3: (a) Unfolded geometry: Light blue dots indicate all 3 000 sensing locations, red dots indicate available sensing locations, and black dots indicate the most informative chosen positions. (b) Exclude positions that are affected by edge effects are excluded using the kNN method. When the center of k-NN positions (gray) is outside the red-dashed disc with a certain radius θ, the position is considered as at the edge (blue), otherwise labeled as feasible assembly position (red). (c) The compressive sensing method is used to reconstruct the original data from a small subset of informative sensor positions.

In the end, we validate our design with the optimal selected sensor positions and the inference model on a real robotic limb (Section 2.3).

2.2.1 Finite Element Simulation

We use a finite element simulation of the 3D structure to get the deformation patterns caused by impact forces. In the simulation, we can apply forces to every location on the surface and record the deformations of all finely discretized positions on the structure, as illustrated in Fig. 2.2. For the example shin structure with a sensing surface area of 200 mm $\times$ 120 mm, we obtain around 4 000 different contact locations and 3 000 different sensing locations.

2.2.2 Filtering Feasible Sensor Positions

Not all the 3 000 sensor positions allow for physical sensor placement due to constraints imposed by the sensor size and sensing range. Strain gauges need to be placed on relatively flat surfaces rather than curved edges. It is disadvantageous to place them near highly rigid supports where only small deformations occur. As shown in Fig. 2.3a, we discard positions close to the rigid support at the top and the bottom of the structure (visualized in a 2D unfolded surface). We use k Nearest Neighbor (kNN) to eliminate candidate positions at the edges by checking whether the center of mass is outside a certain radius θ (Fig. 2.3b). Remaining positions (2 162) are kept (red points) as feasible sensor positions.

Concerning the efficiency in the optimal selection of sensor positions, we use compressive sensing techniques to further reduce the number of candidates. Compressive sensing [137] is a category of techniques optimized to accurately reconstruct sparse or compressible signals from a limited number of measurements. We use the method of Principal Component Analysis (PCA) with QR-Pivoting [138]. It uses PCA to compute the principal components explaining the variance in decreasing order. The QR-pivoting selects those positions that are most important for the top principal components. Details are elaborated in Section 2.2.4, and Fig. 2.3c displaces the linear reconstruction error (unexplained variance) dependent on the number of selected sensor positions. For 1% error, we can select 407 candidates out of 2162 positions (marked as black dots in Fig. 2.3a).

2.2.3 Contact Position Inference

We want to extract higher-level haptic information based on the deformations measured at each candidate position, such as contact locations and force strengths. In this chapter, we formulate our goal as a regression task that directly maps deformations to contact locations. We use the method of Support Vector Regression (SVR) [114] here. SVR is a subbranch of Support Vector Machines (SVM) [139] that combines the strength of non-parametric techniques with efficient storage requirements of parametrized models. The idea is to non-linearly map the input x (deformations) into a high-dimensional feature space using kernel tricks and then find a hyperplane defined by support vectors with a minimum margin for regressing contact locations.

The mathematical model is

$$F(x,w) = \sum_{i=1}^{m} w_i \cdot g_i(x) + w_0, \tag{2.2}$$

where $g_i(x)$ is a nonlinear feature map, w_0 and w_i are the coefficient parameters. The regression error for each example x_i is $\xi_i = \max(0, |F(x_i,w) - p_i| - \varepsilon)$, i. e., the deviation from the target p_i larger than ε. More specifically, the input x is the vector of deformations at the sensor positions, and the target p is the p_x, p_y, p_z position of the contact point. The loss function of the SVR model is formulated as

$$L = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{n} \xi_i, \tag{2.3}$$

where C controls the trade-off between the complexity of the regression model ($L2$ norm of w) and the marginal error. SVR computes scalar products of elements in the feature space as a similarity check between inputs x, which can be expressed as the scalar product using an appropriate kernel function $k(x_i,x_j) = g(x_i)^\top g(x_j)$. We choose the radial basis

function (RBF)

$$k(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \tag{2.4}$$

with hyperparameter γ controlling the sensitivity to distance. We use the package of Python *sklearn* [140] and implement k-fold cross-validation to select the optimal C, ε, and γ . Using all candidate sensor positions and 85% of the data as training data, SVR can achieve an average test error of 0.6 mm ($C = 0.1$, $\varepsilon = 10^{-4}$, $\gamma = 2 \times 10^{-3}$, five-fold cross-validation).

2.2.4 Optimal Sensor Placement

We propose different ways to select the optimal sensor positions, which can be grouped into data-driven methods and model-based methods, as summarized in Fig. 2.4. Data-driven methods rely on previously collected data and select the subset of sensors the regression model performs the best. Model-based methods rely only on the geometric position of the sensors, without dependencies on the collected data. We propose a greedy SVR method since sufficient data is available and the actual dependency between the sensor locations can be accessed. In comparison, we also provide results using a linear compressive sensing method and two model-based methods.

Data-driven method: Nonlinear Greedy Support Vector Regression

The goal is to select a combination of K sensors that perform best on average at inferring the contact position for unseen stimulation. It is empirically intractable to search through $\binom{n}{K}$ different possibilities with $K > 4$. Thus, we use a greedy strategy. We first select the best single sensor position using k-fold cross-validation and then add the second sensor position that gives the best performance and so forth, where the performance is defined in terms of k-fold cross-validation (Alg. 1 and Fig. 2.4).

Data-driven method: Linear PCA with QR Pivoting

As mentioned in Section 2.2.2, the sensor number can be reduced using compressive sensing methods. Here, we provide details about the specific method, i. e., PCA with QR-Pivoting [138]. Compressive sensing relies on two fundamental functional matrices: feature transform basis $\Psi \in \mathbb{R}^{N \times N}$ and sub-sampling matrix $\Phi \in \mathbb{R}^{m \times N}$. Ψ transforms raw measurements $x \in \mathbb{R}^{N \times 1}$ into a sparse representative space $\alpha \in \mathbb{R}^{N \times 1}$ where α has s nonzero elements:

$$x = \Psi \cdot \alpha. \tag{2.5}$$

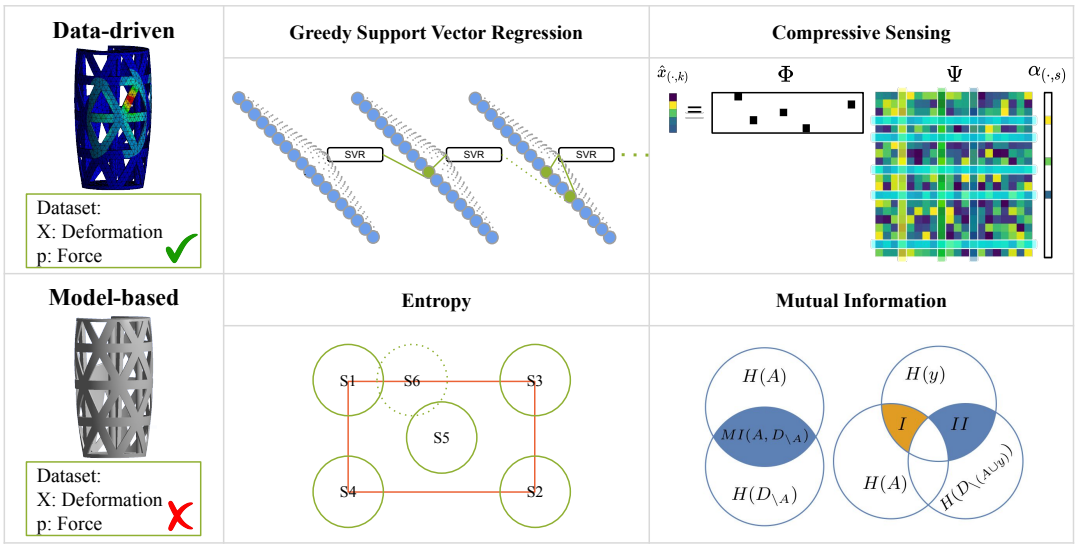

Figure 2.4: Methods for optimal sensor placement. The top row indicates two data-driven methods. Greedy SVR adds the best sensor positions one by one, taking the inference accuracy as the baseline. PCA-QR pivoting takes the reconstruction accuracy of deformation as the target. The bottom row shows two model-based methods. Entropy criterion: Each sensor has the same ability to detect uncertainty indicated as the same radius. Disk S1, S2, S3, S4 are first selected as the overlapping area is zero such that the ability to discover total uncertainty is maximized. However, those areas outside the red FOV are wasted. Mutual Information criterion: The blue shaded area indicates the shared information between the selected set A and the unselected set $D_{\setminus A}$. An explanation for Eq. 2.13: For each step, we choose the sensor y which shares the least information with the set A but represents most information of the rest unselected set $D_{\setminus(A\cup y)}$ and generally maximize area $(II - I)$.

Φ chooses m measurements from N optimally such that the representation α can be most accurately reconstructed from the measurements $\hat{x} \in \mathbb{R}^{m\times 1}$ using the l_1 norm:

$$\alpha = \arg\min_{\alpha'} \|\alpha'\|_1 \qquad \text{s.t.} \quad \hat{x} = \Phi \cdot \Psi \cdot \alpha, \tag{2.6}$$

where the number of measurements and the sparsity of transformed signal have to follow the Restricted Isometry Property [138]: $m \gtrsim s \cdot log(N)$. We transfer the optimal placement task into a linear optimization problem that is conditioned on the operator $(\Phi \cdot \Psi)$. The central challenge is now to find an optimal Ψ compressing raw data x efficiently and design a good Φ such that the operator $(\Phi \cdot \Psi)$ is well-conditioned.

PCA is an unsupervised linear dimension reduction method [141]. It detects the directions of maximum variances in high-dimensional data and projects data onto these directions, forming a smaller dimensional subspace while remaining most of the information. We preserve the first s principal components Ψ_s of Ψ to ensure the s sparsity in α and then choose an optimal sub-sampling Φ to constrain the reconstruction error of

Algorithm 1 Greedy SVR

Input data: Deformation: $X \in \mathbb{R}^{M\times N}$, force position : $y \in \mathbb{R}^{M\times 3}$, maximum sensor budget K
Data standardization & k-fold cross-validation dataset preparation $\{X^1,...,X^k\}$, $\{y^1,...,y^k\}$
Selected nodes: $A = \emptyset$
for $i : 1$ **to** K **do**
  for $m : 1$ **to** $M \notin A$ **do**
    for $j : 1$ **to** k **do**
      $A' \leftarrow A \cup m$
      $e_j \leftarrow \|\mathrm{SVR}(X^j\{A'\}) - y^j\|$
    end for
    $error_m \leftarrow \mathrm{mean}(e)$
  end for
  $A \leftarrow A \cup \arg\min(error)$
end for

Eq. 2.6 based on the condition number criterion:

$$c = \frac{\sigma_{max}(\Phi \cdot \Psi_s)}{\sigma_{min}(\Phi \cdot \Psi_s)}. \tag{2.7}$$

Since Φ is, in fact, a permutation measurement matrix, it can be designed as a column pivoting matrix of Ψ_s. QR factorization with column pivoting selects the highest singular values to maximize σ_{min} and minimize c. Details are shown in Alg. 2 and Fig. 2.4.

Algorithm 2 PCA with QR Pivoting

Input data: Deformation Pattern: $X^{M\times N}$, Maximum Sensor Budget K
Data standardization
Principal Components Decomposition $\Psi^{N\times N}$
for $i : 1$ **to** K **do**
  Pick $1 : i^{th}$ principal components $\Psi^{N\times i}$
  $P \leftarrow$ **QR-Pivoting**$(\Psi^{N\times i})$
  Sensor node marker = $P[0 : i]$
end for

Model-based method: Gaussian Process

Methods that rely only on the geometric location of the sensors can also tackle optimal sensor placement. They are categorized into model-based methods with model assumptions of homogeneous geometry and fixed sensing radius for all sensors. A Gaussian

Process (GP) [115, 117] is such a convenient model. It constructs a covariance structure to measure the similarity between sensor locations, ensuring similar locations with similar sensor values.

A GP procedure specified for our setting can be formalized as in the following. Given a set of sensors A, their positions p_A, and their deformation data $X_{A,\cdot}$, we can predict the distribution of deformations at a different sensor y with location p_y. The mean $\mu_{y|A}$ and standard deviation $\sigma_{y|A}$ are given by

$$\mu_{y|A} = \mu_y + \Sigma_{y|A} \cdot \Sigma_{AA}^{-1} \cdot (X_{A,\cdot} - \mu_A)\,, \tag{2.8}$$

$$\sigma_{y|A}^2 = k(y,y) - \Sigma_{y|A} \cdot \Sigma_{AA}^{-1} \cdot \Sigma_{y|A}^T\,, \tag{2.9}$$

where $\mu_A = \frac{1}{|A|}\sum_{a\in A} X_{\cdot,A}$ is the vector of mean sensor values for the sensors in A, $\Sigma_{AA}(i,j) = k(p_i,p_j) + \beta^{-1}\delta_{i,j}$ is the covariance matrix for all sensors $i,j \in A$ with δ_{ij} being the Kronecker delta, and β is a hyperparameter. Similarly, $\Sigma_{y|A}(i) = k(p_i,p_y)$ is the vector of similarities with the new sensor location.

The GP is a non-parametric process but relies on a fixed kernel $k(\cdot,\cdot)$. Kernels can be polynomial, quadratic, and exponential. In our setting, the deformations can be described as bending a thin plate [142]. Thus, we use an exponential kernel

$$k(x,x') = \exp\left(-\left(\frac{d(x,x')}{l_{scale}}\right)^{l_p}\right) \tag{2.10}$$

with hyperparameters of the length scale l_{scale} and the distance norm l_p. The distance $d(x,x')$ is measured as the approximate geodesic distance rather than the Euclidean distance because the surface of the 3D structure is curved. We use cross-validation [116] to select the hyperparameters. We decide $l_{scale} = 0.033$ and $l_p = 1.9$ based on a grid search.

Probabilistic modeling, here GP, allows us to use information criteria for picking the most informative sensor positions. We evaluated two information-related methods. One minimizes the uncertainty about the non-measured locations (*entropy*) over the whole sensing surface. The other maximizes the *mutual information* (MI) between the selected sensor locations and the non-measured locations.

Algorithm 3 Entropy Criterion

1: Input data: GP, sensor budget K
2: **for** i : 1 **to** K **do**
3: $\quad y^* \leftarrow \text{argmax}_{y\in V\setminus A}\, \sigma_{y|A}^2$ using Eq. 2.9
4: $\quad A \leftarrow A \cup y^*$
5: **end for**

Entropy Sensor locations should be chosen that minimize the uncertainty about the entire permissible locations V. It can be quantified as the conditional entropy of the unobserved locations $V_{\setminus A}$ given the observed ones, i. e., $H(V_{\setminus A}|A)$ [143]. Mathematically, we seek for:

$$A^* = \arg\min_{A \subset V} H(V_{\setminus A}|A) = \arg\max_{A \subset V} H(A)\,, \tag{2.11}$$

as detailed in [143]. Since Eq. 2.11 involves a combinatorial search, we solve it greedily, as implemented in the SVR case (Section 2.2.4). The entropy of a Gaussian distribution is given as $H(\mathbb{N}(\mu,\sigma^2)) = \frac{1}{2}\ln(2\pi e\sigma^2)$. The algorithm is shown in Alg. 3. It suggests choosing sensors far away from each other, as indicated in Fig. 2.4. The selected locations tend to sit on the boundary of the space, inefficiently using their entire detection disk.

Mutual Information Another method used in [143] is Mutual Information (MI). It measures the shared information between selected and unselected locations:

$$\mathrm{MI}(A, D_{\setminus A}) = H(D_{\setminus A}) - H(D_{\setminus A}|A), \tag{2.12}$$

where D is the set of all locations, including not permissive sensor locations (light blue dots in Fig. 2.3a). A graphical explanation is shown in Fig. 2.4. Since maximizing the MI between $D_{\setminus A}$ and A is a combinatorial problem, a greedy method is used. We select the location with maximum additive MI:

$$y^* = \arg\max_{y \in V_{\setminus A}} \left[\mathrm{MI}(A \cup y, D_{\setminus A \cup y}) - \mathrm{MI}(A, D_{\setminus A})\right] \tag{2.13}$$

as detailed in Alg. 4 and intuitively indicated in Fig. 2.4. For each step, we choose the sensor y that shares the least information with set A but represents most of the rest unselected set $D_{\setminus A \cup y}$ and generally maximizes area ($II - I$).

Algorithm 4 Mutual Information Criterion

1: Input data: GP, sensor Budget K
2: **for** $i : 1$ **to** K **do**
3: $\quad B(y) = D_{\setminus (A \cup y)}$
4: $\quad y^* \leftarrow \arg\max_{y \in V_{\setminus A}} \left(\frac{\sigma^2_{y|A}}{\sigma^2_{y|B(y)}}\right)$ using Eq. 2.9
5: $\quad A \leftarrow A \cup y^*$
6: **end for**

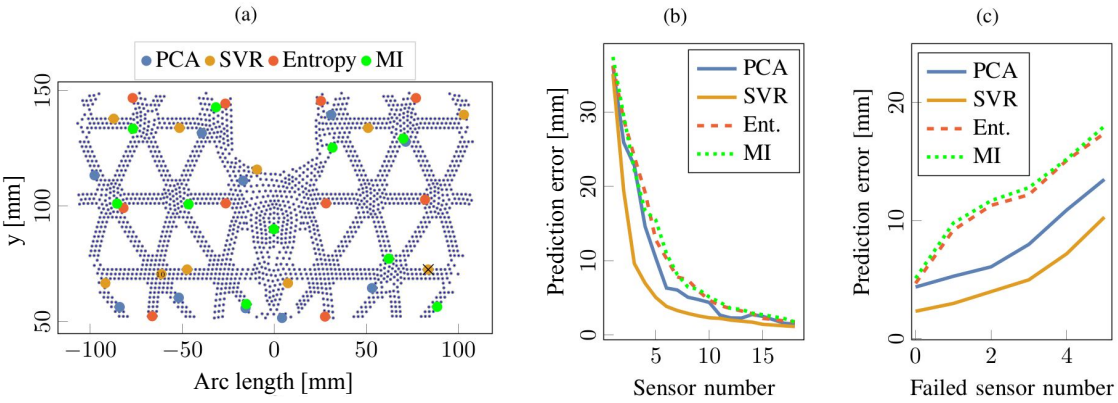

Figure 2.5: Results of optimal sensor placement. (a) Picked optimal positions based on four different criteria. × indicates a physical sensor with failure (Section 2.3.3). (b) Prediction error on test dataset comparing four different methods. (c) Robustness test.

2.3 Results

We first compare the above methods to decide the optimal sensor placement based on the simulation results and then validate it to an actual robotic limb.

2.3.1 Optimal Sensor Placement in Simulation

Here, we compare the performance of the sensor placement using the above methods. We present the selected sensor positions in Fig. 2.5a. The PCA-QR method suggests locations on the edges of each beam whose deformations have the highest variances. The SVR method recommends locations on the two off-center parallel beams ($y \approx 70\,\text{mm}$ and $y \approx 130\,\text{mm}$), where deformation from both sides can be measured. Notice, the 10^{th} sensor position suggested by SVR locates closely to two already selected sensors, while a more centered position would be expected. Further sensors ($11^{th} - 30^{th}$ not shown) recommended by SVR are mainly at the boundaries rather than the center. We hypothesize that the areas near boundaries are more rigid and less sensitive to applied force, which requires a higher sensor density to get high inference accuracy. The model-based methods (Entropy, MI) are purely based on geometric information. The Entropy criterion suggests placing sensors toward the edges and homogeneously distribute them on the entire space. The MI criterion recommends that the positions be more centered.

After selecting the optimal sensor positions, we compare the four methods using the SVR contact location inference to evaluate the prediction performance(Section 2.2.3). As indicated in Fig. 2.5b, data-driven methods work better because they can exploit the data structure. The PCA-QR is not a greedy method and suggests different combinations of sensors for each sensor budget K. Thus, the prediction error is not guaranteed to be in descending order. In conclusion, a few sensors can already lead to a small prediction error.

Table 2.1: Strain gauge assembly procedure

Step	Details
1	Wire SG and cover SG with scotch tape to isolate adhesive.
2	Cover SS with preservative film to isolate adhesive.
3	Insert absorbent wool between SG and SS to absorb the adhesive.
4	Position SG on SS.
5	Clean internal skeleton surface and SG surface.
6	Coat SG and internal skeleton surface with prepared adhesive.
7	Pre-tighten the whole structure and cure for $6\,h$ at $+20^{\circ}C$.
8	Disassemble SS and clean surface.

Using the simulation data, we acquire a precision $< 10\,\text{mm}$ for five sensors, and roughly 2.5 mm for ten sensors, selected using the greedy SVR method (Section 2.2.4).

Moreover, we also check the robustness of the selected positions against failing sensors, as any physical device is susceptible to failure. We test the prediction error with different degrees of sensor failures, i. e., out of ten sensors, 1 to 5 are broken. As shown in Fig. 2.5c, the greedy SVR method has the highest robustness against sensor failures. For $3/10$ failures, the performance drops by 5 mm.

2.3.2 Hardware Setup

We use the ten sensor positions suggested by the greedy SVR method and implement it on an actual robotic limb. In the following, we introduce the hardware setup details that include the physical sensor selection, the assembly process, and the data acquisition procedure.

Sensor choice We use strain gauge (SG) sensors because they are generally affordable, widely available, and relatively straightforward to use. For the 3D-printed plastic robot parts with potential strong deformations, we choose SGs with a 20% elongation rate. It ensures that the SG's finite extension is within the maximum elastic deformation of the limb. The chosen SG (Micro Measurement, EP-08-250BF-350) has a long lifetime and high fatigue strength, while it can only measure deformation along one direction. The SGs are suggested to be assembled along with the beam directions according to finite element simulations.

Sensor assembly Strain gauges have to be attached to the inner side of the 3D plastic structure with precaution to avoid damage or malfunction. The assembly procedure is challenging because it is inside the hollow structure. We develop an assembly approach with a custom support structure (SS), as shown in Fig. 2.6a. The support structure has little

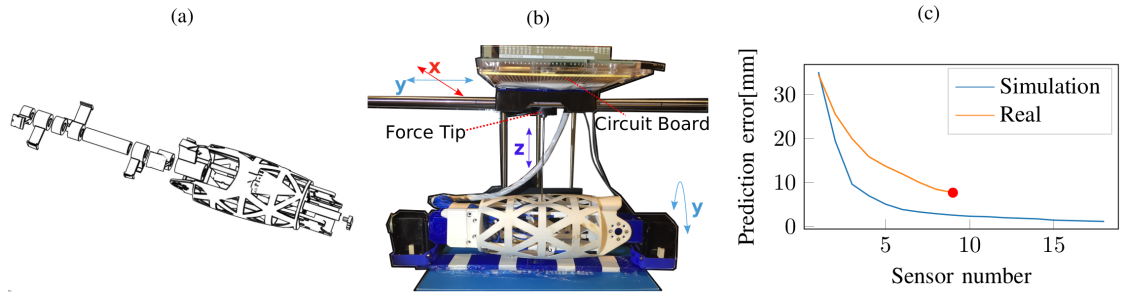

Figure 2.6: Experimental setup and results. (a) Support structure (SS) for assembling strain gauges (SG); (b) Testbed with four DoFs to collect data of single contact on the surface; (c) Prediction error comparison between simulation and real physical setup.

arms to pre-tighten each SG at the selected position. The arms are held by a middle axis during the adhesive curing process and can be pulled out for disassembly. The assembly process is presented in Tab. 2.1.

Data acquisition To acquire the dataset for training the machine learning algorithm, we record sensor measurements for different force stimulations. The classic data acquisition circuit for SG is a Wheatstone bridge [144] that measures electrical resistance change by balancing two legs of a bridge circuit. We integrate a temperature compensation function into the circuit because SGs are sensitive to mechanical stress and temperature variance. We use a half-bridge Wheatstone layout, an operational amplifier of MCP609, and an Arduino Due. The circuit has twelve ADC I/O ports with 12 bits of resolution. The SG's deformation is amplified by a factor of 330 and converted to 4096 different values over 3.3 V. We build a testbed with a force sensor tip to collect numerous sensor-force measurements in an automated way. The testbed offers three Degrees of Freedom (DoF) in translation ($\vec{x}, \vec{y}, \vec{z}$) and one DoF in rotation ($\circlearrowright \vec{y}$). The testbed is equipped with a force sensor (FC2231). The whole setup is visualized in Fig. 2.6b.

2.3.3 Experimental Results

We validate the proposed "HapDef" system on a modified limb of the Poppy robot [135]. The limb has rigid internal support and a flexible shell to detect contacts. The support structure sustains the forces acting at the joints and avoids shell deformations under no contact. We assemble ten SGs to positions suggested in Fig. 2.5a. One of the SGs is malfunctioning, which is shown as a cross in the figure. Each SG value is calibrated to be zero if there is no contact.

We visually discretize the outer surface of the flexible sensing shell into homogeneously distributed points (3 000), avoiding the edges and boundaries so that the force-tip does not slide off. Then we let the testbed automatically move the force-tip towards the flexible

Table 2.2: Inference error of contact position and magnitude w.r.t. force strength

Force Interval [N]	Position Error [mm]	Amplitude Error [N]
0 - 4.9	25.43 ± 13.36	1.05 ± 1.01
4.9 - 9.8	11.95 ± 11.85	1.19 ± 1.19
9.8 - 19.6	5.90 ± 7.79	1.42 ± 1.78
19.6 - 34.3	4.48 ± 6.29	1.54 ± 2.21

shell to apply normal forces at each location. Upon contact establishes, the testbed moves the force-tip with incremental steps of 0.2 mm to a maximum penetration depth of 2 mm.

We regroup a dataset by choosing the largest force simulation at each contact location (3 000 locations in total) and split the dataset into training/validation/test subsets with a ratio of 8/1/1. We train an SVR model by performing hyper-parameter selection based on the validation set ($C = 20$, $\varepsilon = 10^{-6}$, and $\gamma = 2 \times 10^{-3}$). We evaluate the inference performance on the unseen contact locations. Under consistent conditions, we implement the above procedures also in a simulation setup. As shown in Fig. 2.6c, the hardware implementation achieves half of the simulation accuracy. The average inference accuracy of the contact position is below 8 mm over the whole surface of 200 mm $\times$ 120 mm using only nine physical sensors.

Moreover, we train an SVR to evaluate the inference accuracy w.r.t. contact strengths, varying from light to strong. As shown in Tab. 2.2, SVR has low inference accuracy for light contact and high for strong because fewer sensors get activated by light stimulation. The absolute inference accuracy of force strengths varies little w.r.t. force strength. Consequently, any strong touches on the surface can be reliably detected and be used as a warning signal, which improves the haptic system's robustness.

2.4 Conclusion

We introduce a method to obtain a robust haptic sensing system using only a few deformation sensors. A machine learning approach powers the sensor device. After a learning period, the system can reliably localize contact all over a curved surface. Moreover, the system is very durable as the physical sensors are placed inside the structure. Only nine sensor values need to be acquired and processed. The computational requirements are also low during operation, as the inference of the force location is made via Support Vector Regression. However, other machine learning methods, such as Deep Neural Networks, are also feasible. In addition, we evaluate different methods of choosing optimal sensor locations for a sparse sensor configuration. We find that data-driven methods outperform geometry-based methods. The proper selection strategy can reduce the required number of sensors by 50% without significant loss in accuracy. In the next chapter, we will investigate more accurate force information inference and multiple-contact detection.

Chapter 3

Transfer Learning for Multiple-Contact Inference

This chapter introduces a general real2sim machine-learning framework to infer multiple-contact haptic information for the modified "Poppy" limb stated in Chapter 2. This framework solves the data insufficiency problem for multiple-contact scenarios, where we can only trivially collect single-contact data in the real physical world out of complexity consideration. The general idea of this framework is first to build a finite element simulation model to describe the mechanical behavior of multiple contacts and reformulate the model as an end-to-end machine learning model. Then we calibrate the simulation model with the real sensor system using transfer learning based on the real single-contact data. We validate the framework performance to improve the single-contact accuracy up to 3 mm and obtain 10 mm accuracy for double-contact.

3.1 Introduction

Haptic feedback at large parts of the body is essential for robots to learn interaction patterns, exploit the environment, and detect unexpected or safety-relevant situations for mastering real-world challenges. Haptic systems for large-scale sensing should ideally provide contact parameters, such as contact location and force information for multiple contacts, with high spatial and temporal resolution (referring to human sensing [145]). In addition, such systems should be: robust to long-lasting impacts, low-cost, energy-saving, and computationally inexpensive.

The work done in Chapter 2 shows advantages in large-scale sensing with a trade-off of the above-listed criteria, which employs machine learning for single-contact estimation from a sparse sensor configuration. With this method, contact position and force magnitude can be inferred with sufficient accuracy on a robot limb with a surface of about 200 $\times$ 120 mm equipped with only nine strain gauges (8 $\times$ 5 mm). Using the same physical setup and taking it as a basis, we explore in this chapter the potential of the setting for more precise estimation and the extension to multiple-contact scenarios (Fig. 3.1).

A data insufficiency problem arises for multiple-contact detection within the data-driven inference scheme. Because performing data collection for multiple contacts is

Figure 3.1: Ultimate goal: Inferring multiple-contact haptic stimulation from a few measurement elements.

physically challenging. Multiple independent force-tips operated by more than one robot arm would be needed to stimulate the haptic system in a controlled way. Thus, in this chapter, we propose a real2sim framework to solve this problem. We build a finite element simulation model to analyze the mechanical behavior of multiple contacts, collect data for multiple-contact cases and calibrate the simulation model with the real system using physically collected single-contact data. In this way, the knowledge of multiple contacts is transferred from the simulation model to the real physical setup. We name this procedure as a real-to-simulation transfer learning framework (shortened as real2sim framework).

To evaluate the multiple-contact tactile spatial accuracy, we refer to the acuity of the human tactile perception threshold quantified by the "two-point discrimination" criterion, widely adopted to examine tactile perception in clinical settings [146, 147]. It indicates the ability to discriminate two nearby stimulation on the skin into two distinct contacts instead of one. In human bodies, this ability essentially differs from body part to body part [145]. We will compare our real2sim framework performance with the acuity on the shin, palm, and fingertip.

This chapter is structured as follows: Section 3.2 presents the method by first giving an overview and then investigating the different aspects from single- to multiple-contact prediction. In Section 3.3, we present the results on the robotic limb. We close with a discussion in Section 3.4.

3.2 Methods

To tackle the multiple-contact estimation task, we propose a new pipeline based on neural networks. We begin with the problem description. Given readings of a few physical

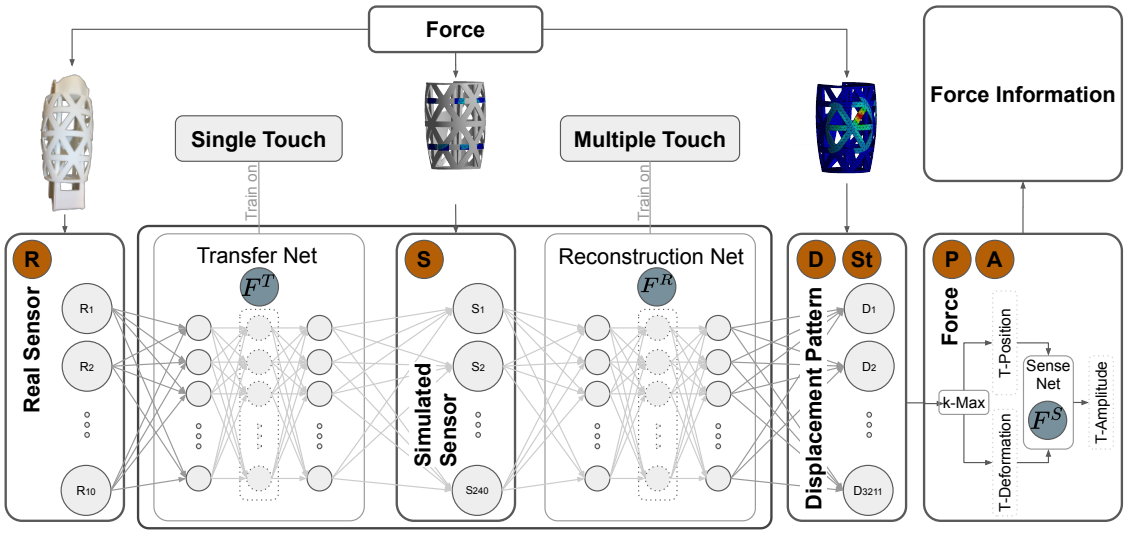

Figure 3.2: The general framework for multiple-contact haptic information estimation: The framework first calibrates the simulated sensor values with the ten real physical strain gauge values using **Transfer Net** F^T and then predicts the displacement pattern based on the FEM simulated sensor values using **Reconstruction Net** F^R. Based on the predicted displacement pattern, we extract the T candidate contact locations exploiting the **k-Max** method. Relying on the **rigidness distribution** of the 3D structure, we infer the force strength of each candidate contact according to the location and the displacement using **Sensitivity Net** F^S. All inferred candidate contact information is then summarized into the robot system for further post-processing to rule out the spurious contacts.

sensors $R \in \mathbb{R}^{N_R}$ ($N_R = 10$), we infer two quantities to describe the multiple-contact information. A displacement map of the entire surface is indicated by the displacements at many points (nodes) $D \in \mathbb{R}^{N_D}$ ($N_D = 3211$), with $D_i = 0$ denoting no displacement at a particular point. The displacement map is similar to a visual input indicating the interaction forces by a pattern. Another quantity is the explicit contact point information, i. e., the positions $P_i \in \mathbb{R}^3$ and their respective force impact $A_i \in \mathbb{R}$ for all the unknown T contact points $i \in [1, \ldots, T]$; here, we only consider forces in the normal direction. This quantity is with lower dimensions and allows for a direct quantitative comparison in terms of position and force (magnitude) accuracy while being more difficult to obtain compared to the pure displacement map acquisition.

As presented in Fig. 3.2, the pipeline comprises three sub-tasks: First is transfer learning from real to simulated sensors (Transfer Net F^T); the second is displacement pattern reconstruction from simulated sensors (Reconstruction Net F^R); the third is the position and force magnitude estimation (position and number of contacts detector *k-Max* and Sensitivity Net F^S). Each neural network is trained with its own training scheme and dataset, which we explain in the following.

3.2.1 Transfer Learning

We implement the transfer learning procedure (**Transfer Net** F^T), mapping real sensor values R to simulated sensor values S. We design a four DoFs testbed equipped with a force-tip to collect single-contact data. The data contains real sensor, and corresponding force contact positions and strengths all over the 3D surface. As done with the testbed, we build a finite element simulation model for the robotic limb structure to obtain the simulated sensor values under consistent contact situations. The real sensors (strain gauge) measure the surface elongation when the surface bends due to externally applied force. The simulated sensor values are computed by a FEM tool ANSYS [136] and are delivered as individual point's displacement. To align the real and simulated sensor measurements, we assign a "virtual sensor" comprising a patch of points around each physical sensor, as shown in Fig. 3.5B. We have a patch of 24 virtual points for each strain gauge in our implementation, as explained in the following Section 3.3.2. The target values for the transfer net F^T, i. e., the simulated virtual sensor values, are obtained by applying the same forces in the simulation. The net is trained by minimizing the mean squared error:

$$L^T(\theta) = \mathbb{E}\left[\|F_\theta^T(R) - S\|^2\right]. \tag{3.1}$$

3.2.2 Reconstruction of the Displacement Map

We use the **Reconstruction Net** F^R mapping S to D to infer the whole displacement map, i. e., the displacements at all points on the surface. The data for training this network is collected solely by simulations for different numbers of contact points and force strengths. The training procedure is with a minimization objective of the mean square loss:

$$L^R(\theta) = \mathbb{E}\left[\|F_\theta^R(S) - D\|^2\right]. \tag{3.2}$$

3.2.3 Force Position and Magnitude Estimation

We can extract the number and the position of the interaction points from the inferred displacement pattern. Intuitively, we assume the displacement to be locally maximal for the points where the forces apply. It is consistent with studies on thin-plate displacements [142], where the displacement at the contact force location is the highest and decreases with the distance approximately following a bell shape. Thus, we implement a simple k-nearest neighbors algorithm called **k-Max** to detect the local maxima. A candidate force location is chosen when the location is with the largest displacement among its k nearest neighbors. Our implementation extracts only one maximum by specifying a spatial region around a 10 mm radius as the neighbors. The approach tends to find more contact points than there are in reality due to non-linearity in surface displacements.

The **rigidity of the surface** varies across the sensing surface, and we learn the mapping from displacement to force strength (**Sensitivity Net** F^S). It is trained with the squared-

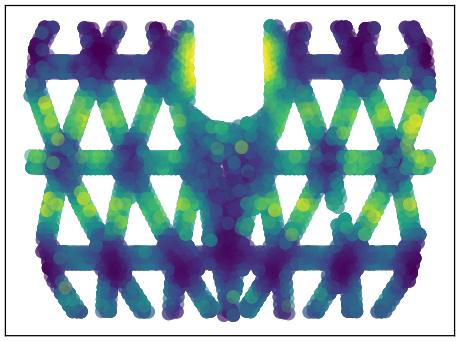

Figure 3.3: Rigidness evaluation of the 3D structure: The same force is applied on each point on the surface, and the displacement is compared. Lighter colors indicate more significant displacements and less rigidity (normalized for visibility).

error loss:

$$L^S(\theta) = \mathbb{E}\left[\|F_\theta^S(D_i) - A\|^2\right], \tag{3.3}$$

where D_i is the displacement at the location of contact, and A corresponds to the respective force (strength). The rigidity of the surface can be visualized as F^S with a constant force input, see Fig. 3.3. The upper and lower areas connected with the thick joint boundaries, and the hexagonal connection hub are the most rigid areas of the surface.

3.2.4 Architectural Choices

Our proposed framework comprises three neural networks. In this section, we investigate some systematic optimization procedures for the architectures.

Real Sensors and Transfer Net The number and the placement of the physical sensors are optimized using a greedy strategy to maximize single-contact performance, as done in Chapter 2. The Transfer Net F^T is a fully connected feed-forward neural network (FNN) (6 layers with 250 tanh hidden units each), and is trained on 27 000 data points (R,S) with a training/validation/test ratio of 3:1:1. The number of layers and the units are chosen through Bayesian hyperparameter optimization.

We can theoretically train each real sensor independently to analyze the generalization property of the relationship between real physical sensors and simulated sensors. However, concerning the inhomogeneous rigidness of the 3D structure, we design the network that connects all real sensors to all simulated sensors. Moreover, the number of simulated sensors affects the generalization performance in the case of multiple contact points. On the one hand, the number can be chosen as high as possible from the perspective of multiple-contact prediction from simulation. The extreme case is that the number of simulated sensor points equals the number of displacement points on the surface. It makes

the reconstruction step unnecessary. On the other hand, we wish the entire framework can generalize from single-contact data during training of the transfer net F^T to multiple-contact without retraining it, because we have no data to do so. It suggests to exploit only a few simulated sensors because the transfer learning should capture the mapping from real sensors to the simulated patches around each sensor individually. It enables the generalization to multiple contact points where the sensors are differently correlated. Since the physical sensor and the simulated sensor measure different quantities, we have chosen a small patch around each physical sensor location reflecting roughly the surface of the physical sensor, as displayed in Fig. 3.5B. The number of simulated points is optimized hyperparameters.

Reconstruction Net The architecture of the Reconstruction Net F^R is an FNN with 600-1200-1800 ReLU hidden units as a result of hyperparameter optimization using validation data of single-, double-, triple-, and quadruple-contact.

Force Position and Magnitude Inference We consider different algorithms like Gaussian Mixture Model (GMM) [148] and fitting Radial Basis Functions (RBF) to find the contact points in the predicted displacement map. However, GMM cannot be applied directly, and the data would need to be transformed since the task is not a density estimation. For fitting the RBFs, a good initialization is needed to obtain consistently good results [149]. Our proposed nearest neighbor approach (**k-Max**) shows superior performance in our setting. The Sensitivity Net F^S is an FNN with four layers, each with 250 hidden ReLU units mapping contact location and displacement to the force strength. It is trained on 27 000 data points (local displacement, position $\rightarrow$ A) with a training/validation/test ratio of 3:1:1.

3.3 Results

We first evaluate the performance of the single-contact inference using neural networks and then investigate the performance of different parts of our architecture related to the multiple-contact generalization. Afterward, the capabilities of the proposed system are tested on up to four contact points.

3.3.1 Direct Single-Contact Inference Baseline

A simple regression model can infer single-contact information, directly mapping real sensor values (R) to contact position and strength (P,A). In Chapter 2, we used support vector regression (SVR). We employ both a k-nearest neighbors (k-NN) regressor and an FNN for this task and summarize the results in Fig. 3.4A. The FNN outperforms k-NN and SVR with a small margin. k-NN and SVR predict each coordinate independently, while the FNN makes joined predictions, exploiting correlations in the output space. Fig. 3.4B

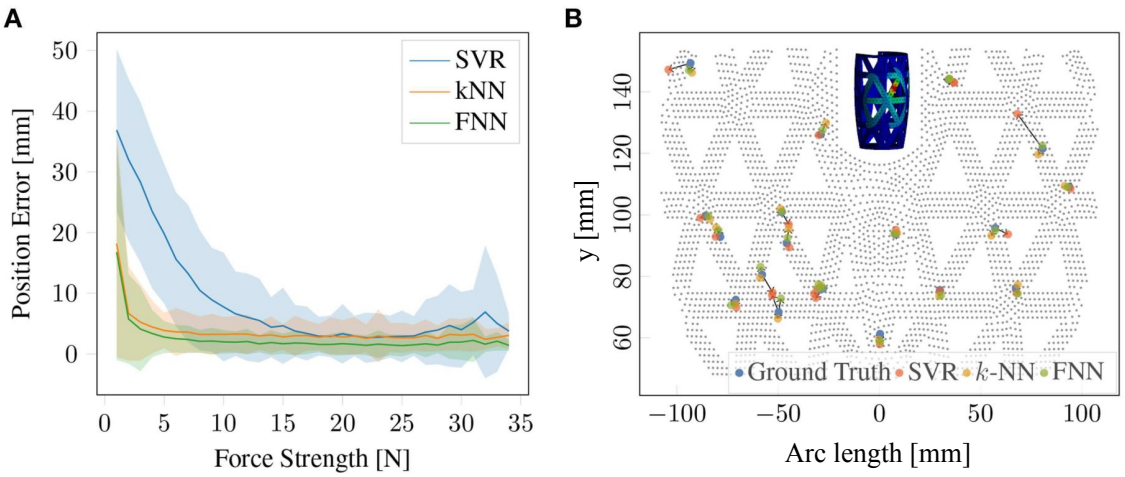

Figure 3.4: Single-contact information inference baseline: **A** Inference accuracy of contact position using SVR, k-NN, and FNN. **B** Comparing predictions of SVR, k-NN, and FNN for the single-contact position on a 2D surface projection. Gray dots represent the geometry grid, blue dots are the ground truth contact positions, red, orange, and green dots are the predicted contact positions using SVR, k-NN, FNN, respectively. Arrows are the error vectors.

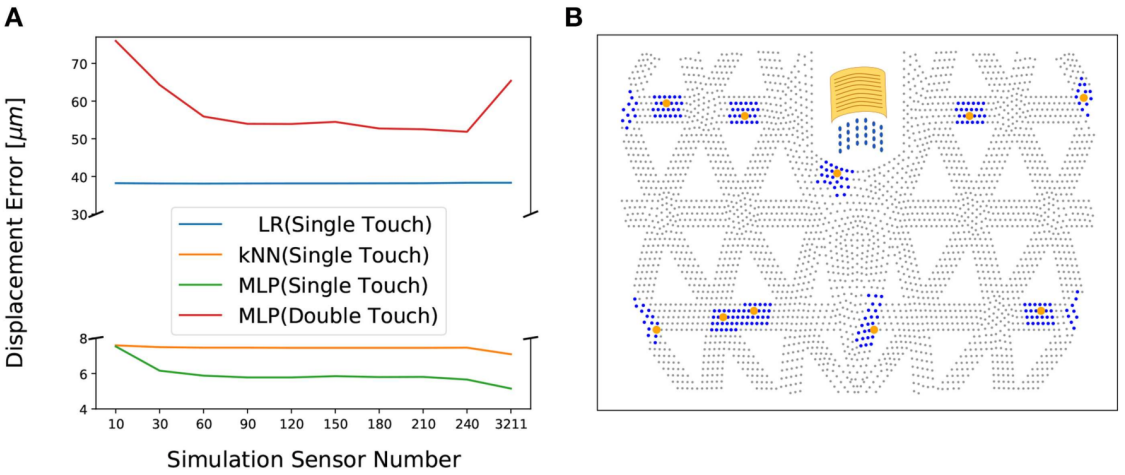

Figure 3.5: Transfer learning maps the physical sensor values to simulated sensor values: **A** The effect of the simulated sensor number on the inference accuracy (of the full displacement map) is evaluated with three methods: Linear regression (LR), k-NN, and FNN (F^T). The training is implemented on a large set of single-contact data, and the reported test performance is for single-contact and double-contact data collected on the real system. The reconstruction net F^R is trained on the simulated multiple-contact data. Twenty-four simulated sensor points per physical sensor, yielding 240 sensors, are optimal for generalizing to double contact prediction. **B** The orange dots are the centers of physical SGs (optimally selected in Chapter 2, and the yellow sheet illustrates the real physical strain gauge. The blue dots around each strain gauge's center are the simulated sensor points (nodes in the finite element simulation model).

shows the predictions for random contact positions and force strengths. In summary, the FNN model is a comparably suitable architecture for the regression problem and is used as an optimal baseline for the single-contact inference.

The hyperparameters for three methods are chosen based on 27 000 training and validation data samples, split as above. SVR: $C = 20$, $\varepsilon = 10^{-6}$, and $\gamma = 2^{-3}$; k-NN: $k = 6$ nearest neighbors, weighted by distance; FNN: five layers with 500 ReLU units each.

3.3.2 Generalization of Displacement Inference on Multiple-contact

We employ a transfer learning approach to infer multiple-contact haptic information without requiring massive data samples from the real system. As shown in Fig. 3.2, the real sensor values can be correlated to simulated sensors. We can train a model from simulated sensors to a whole displacement map using the FEM simulated multiple-contact data. One of the design choices is to decide how many simulated sensors around each physical sensor should be used. In Fig. 3.5A, we evaluate single- and double-contact prediction performance depending on the number of simulated sensors. We train the transfer learning part using single-contact data only (27 000 points) and evaluate the test performance based on the real system's single- and double-contact data. The optimal number is chosen to be twenty-four simulated sensors per strain gauge (resulting in 240 simulated sensors), yielding the best performance on double-contact. The arrangement is around the center of the physical SGs, as indicated in Fig. 3.5B.

With the fixed simulated sensor arrangement, we evaluate the capabilities of the reconstruction network F^R for inferring the displacement map. We first generate data in simulation for different numbers of contact points stimulating the outer surface. We collect single-, double-, triple-, and quadruple-contact with randomly selected contact positions and force strengths (each type with 27k, 6k, 6k, and 6k samples with a training/-validation/test ratio of 3:1:1). We then evaluate the inference performance of different models for the reconstruction process, namely linear regression (LR), k-NN, and FNN (F^R). Table 3.1 summarizes the results for the accuracy of predicting the global displacement map depending on which dataset is used for training and testing. We can extract two important messages: First, linear regression is best at generalization from single-contact to multiple-contact prediction, even though the results are poor. Second, the FNN performs best when the respective number of contact points is present during training. It also generalizes best from single-, double-, and triple-contact data to predict quadruple-contact. Thus, using training data for multiple contact points is beneficial. In the following, we will use the FNN for training F^R with all contact data.

We visualize the comparison between simulated and predicted displacement maps in Fig. 3.6. They show remarkable one-to-one coherence. The predictions for single- and double-contact are based on **real physical sensor values**. Due to hardware limitations, real sensor values for triple- and quadruple-contact are not validated. Instead, we show the results for the reconstruction net F^T using **simulated sensor values** in Fig. 3.6C. We evaluate the method on prediction error of the displacements using the normalized

Table 3.1: Displacement map reconstruction: Comparison of multiple-contact inference methods and datasets for predicting the displacements D from the simulated sensors S. I, II, III, IV stand for datasets of single-, double-, triple-, and quadruple-contact respectively. Grey columns show extrapolation to the unseen number of contact points.

		Displacement Prediction Error [μm]															
LR	Train	I				I & II				I, II & III				I, II, III & IV			
		2.2				6.2				10.9				13.6			
	Test	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
		2.2	36.6	60.3	78.2	1.9	16.1	24.4	31.8	2.0	15.6	23.0	29.8	2.1	15.6	22.6	29.0
k-NN	Train	I				I & II				I, II & III				I, II, III & IV			
		2.3				12.8				27.5				40.9			
	Test	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
		2.3	159.0	225.1	286.9	2.4	36.3	97.8	144.6	2.3	36.8	75.8	109.6	2.3	37.8	73.36	99.8
FNN	Train	I				I & II				I, II & III				I, II, III & IV			
		0.6				2.9				5.7				8.1			
	Test	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
		0.6	82.3	123.6	158.1	**0.7**	**7.7**	30.6	48.6	**0.8**	**7.0**	**15.5**	**25.7**	**1.1**	**6.8**	**13.4**	**20.2**

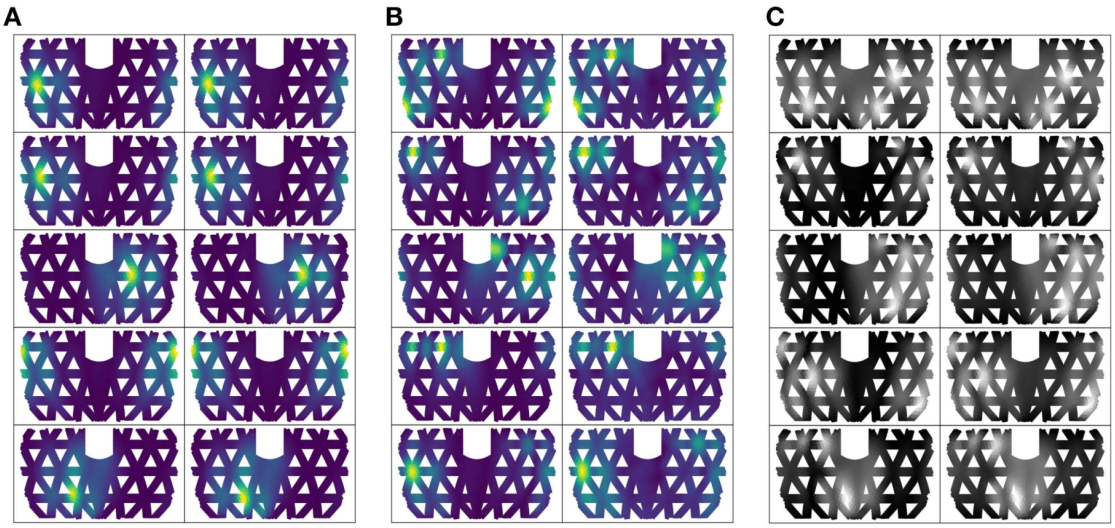

Figure 3.6: Displacement map inference results: **A** and **B** are displacement patterns of single- and double-contact based on the data collected in **the real system**. The left columns show ground truth (simulation: S to D), and the right columns are the predictions (real: R to S to D). **C** presents predicted displacement patterns of triple contacts based on **simulated sensor values** (S to D). R, S, D, P, and A are indexes from Fig. 3.2, indicating real sensor values, simulated sensor values, nodal displacement map, contact locations, and force strengths, respectively.

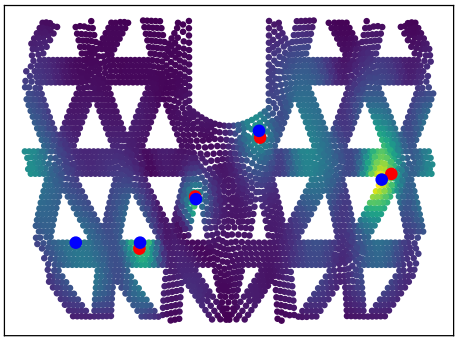

Figure 3.7: Location extraction of quadruple contact. We extract the locations of multiple contacts using the **k-Max** method, based on the predicted displacement map (lighter color indicates stronger displacement). Red dots show the ground truth, and blue dots are the predicted contact locations.

fraction of variance unexplained (FVU). An FVU value of zero corresponds to perfect prediction, and a value of 100% means the inference is as bad as predicting the mean of targets. The FVU shows the prediction of the displacement is very accurate, reaching 4% for single-contact and 29% for double-contact. Nevertheless, these numbers do not explicitly indicate the detection accuracy of individual stimulation. We will investigate this in the following.

3.3.3 Contact Location and Force Magnitude Prediction

We use **k-Max** to extract the locations of the contact points from the predicted displacement maps. As shown in Fig. 3.7, the contact locations for four contact points are extracted. The predicted locations show a small offset w.r.t. the contact positions. Because, for specific geometry, the most significant displacement tends to be on the beam edges.

The localization of single contact using **k-Max** is slightly worse than direct prediction using **FNN** directly, as shown in Fig. 3.8A. We do ablation study for different parts of the architecture. The results illustrate that the Reconstruction Net needed for multiple-contact does not affect the performance of single-contact inference. However, the direct prediction of the contact position and force magnitude is impossible for multiple contact points. For the 3D curved limb structure, we can get candidate predicted contact locations without a corresponding stimulation, shown as blue dots in Fig. 3.7. The spurious ones (the left blue point) typically have the lowest magnitudes among all candidate points.

To predict the contact force strength, we consider the rigidity of the surface, as shown in Fig. 3.3. The same force causes different displacements at different points over the surface. We use the Sensitivity Net F^S to capture this property that considers the position and the displacement at the detected peak displacement and predicts the magnitudes for each potential contact position. The localization accuracy of single-contact force strength using **k-Max** worsens with increased force strength, but the relative error is still low, as shown in Fig. 3.8B. The performance can be improved when the Sensitivity Net F^S is

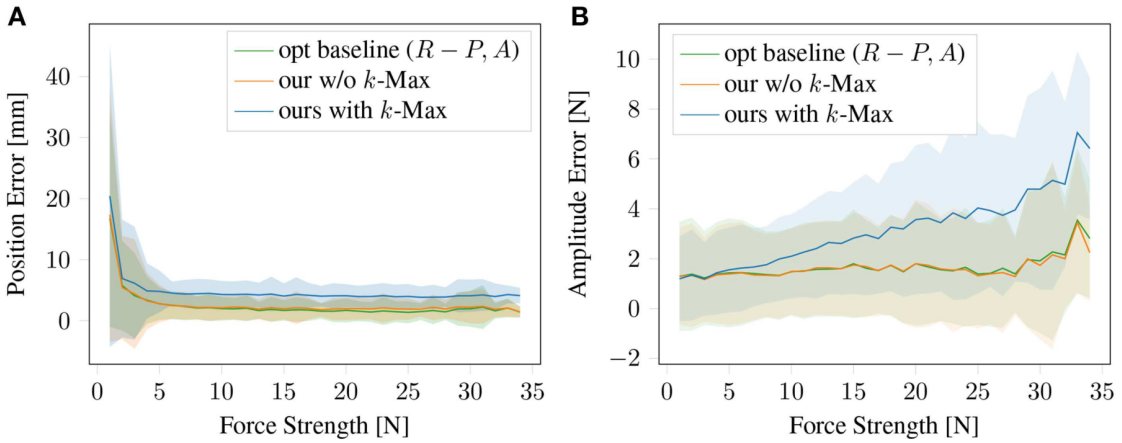

Figure 3.8: Contact position and force strength prediction for the single contact. We compare our architecture with two other methods: the optimal baseline mapping directly from sensors to the target, and our entire framework without **k-Max** but predicting P,A from D directly. **A** shows the contact position prediction error, and **B** presents the force strength prediction error.

trained jointly to compensate for systematic errors in contact-location detection. The performance of the **k-Max** procedure is slightly lower than the direct methods, but it can extract multiple contact points.

3.3.4 Testing the Whole Framework

We evaluate the whole pipeline on single- and double-contact data collected from the real robotic limb and quantify the performance. We collect a double-contact dataset on the real system for 270 pairs (with a distance of 12 mm–75 mm) of contact points with different magnitudes. One contact has a fixed force strength around 30 N (by the robot arm), and the other varies in the range of 0 N–34 N (manually applied with a force-tip).

Our proposed force-extraction method (k-Max and Sensitivity Net F^S) yields a set of potential contact points with respective corresponding force magnitudes. Spurious contact points typically have significantly lower force magnitudes than the actual ones, visible by marker size in Fig. 3.9. For evaluating the distance metric for contact position and strength estimation, we first sort the extracted contact points by their force magnitudes and then calculate the average distance of the top T contact points with the closest real ones. T stands for the number of actual contacts. The average precision for double-contact is 14 mm for the location and 6 N for the force strength when we take the sorted highest T contacts' magnitudes as targets.

Spurious points often lie in between multiple contact points with similar strength. To eliminate the spurious points, we consider evaluating the detected contact point $T+1$. We select it instead of T, if and only if it exists and is within 10% amplitude difference compared to the contact point T. With this implementation, the precision for double-

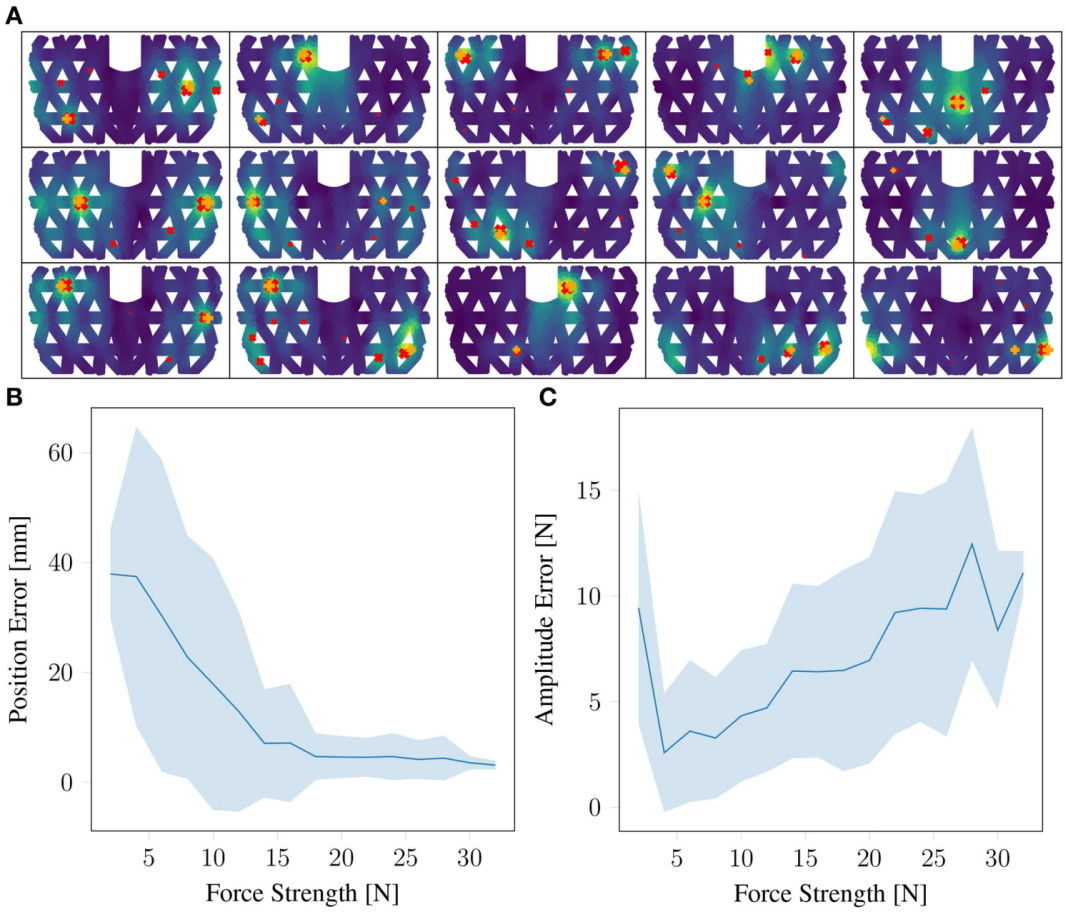

Figure 3.9: Results of the entire framework for double-contact detection: **A** presents the predicted displacement patterns and detected contact points for double-contact (orange $+$: true points; red $\times$: detected). The size of the crosses indicates the force strength. **B** is the prediction error for double-contact w.r.t. the force strength. **C** is the evaluation for the force strength.

contact is optimized to 10 mm for the location and 6 N for the force magnitude, as presented in Table 3.2. Fig. 3.9(B and C) show the error of contact location and force magnitude w.r.t. the applied force strengths.

Our proposed real2sim framework infers force strength information correctly for robotic multiple-contact applications. The haptic sensing system has a higher localization acuity than the human shin and is on par with the human palm. The comparison to the human skin is to give an intuition and is not rigorously conducted. Our system has a hollow structure with holes that are "blind" to touch, but these are not considered in the statistics. Moreover, we compare the absolute localization performance of our system with the bare ability to distinguish two stimuli measured in human subjects.

Table 3.2: Estimation accuracy summary: Comparison of the inference performance of HapDefX (real2sim framework), HapDef (SVR), and Human skin. The bold numbers show accuracy improvements by the methods used in this chapter. The number for human skin is the "two-point discrimination" distances taken from [145].

Precision	Single-contact			Double-contact			
	HapDef	**FNN**	**HapDefX**	**HapDefX**	Human fingertip	Human palm	Human shin
Location [mm]	8	**3**	**5**	**10**	2–8	8–12	30–40
Magnitude [N]	1.6	**1.5**	**3**	**6**	–	–	–

3.4 Conclusion

We present a robust, affordable, large-surface haptic system with sparse sensor configuration capable of sensing contact locations and force strengths for multiple contacts. The system is powered by a machine learning approach, which can reliably localize multiple contacts over a curved surface and estimates the respective force strength. The example shown in this chapter uses only ten strain gauges covering a sensing surface of 200 mm$\times$120 mm to achieve a localization accuracy of 5 mm on single-contact and 10 mm on double-contact. The updating rate of the hardware can be up to 1 kHz using an Arduino DUE, and the FNN processing is also fast due to the use of a small-sized neural network. The proposed method can be applied to any large surface system where a set of sensors can be attached to a deformable surface. This chapter shows that training data on single-contact points is enough to detect multiple contact points for a real physical system. In the next chapter, we validate the proposed real2sim framework on another type of sensor with different physical settings.

Chapter 4

Transfer Learning for an ERT-based Fabric Sensor

In this chapter, we validate our proposed real2sim machine-learning framework for the multiple-contact detection purpose with another sensor. The sensor is made of multiple conductive fabric layers, and it relies on the electrical resistance tomography (ERT) technique. ERT is favorable for fabricating large-scale soft tactile sensors that are flexible and robust. It measures numerous non-local combinations of pairwise conductivity using a relatively small number of electrodes. This measurement method has strong non-locality, which is challenging for transfer learning. A multiple-contact case could be generalized as a combination of single contacts when strong superposition with locality exists in sensor measurements as in Chapter 3.

Here, we aim to investigate the usability of the real2sim framework in estimating multiple contact locations and accurate contact forces for the ERT-based sensor. The real2sim framework comprises two deep neural networks: the first one (Transfer-Net) captures the inevitable gap between simulation and reality, and the second one (Recon-Net) reconstructs contact forces from raw sensor measurements. The sensor has a 540 mm by 560 mm rectangular profile. When tested with 150 unseen multiple-contact cases, the proposed pipeline gives an average contact position error of 74.4 mm (about 13% of the sensor width), where force estimation and contact diameter prediction errors are 17.2% and 9.3%, respectively. These results demonstrate that our approach outperforms both conventional physics-based mapping and direct end-to-end learning in estimating multiple-contact forces from voltage measurements.

4.1 Introduction

Robotic haptic skin should provide scalability to cover sensing areas that are large and often curved, as well as the ability to sense multiple simultaneous contacts. Rather than utilizing many small sensing elements in arrays [150] or modules connected by digital communication [12], reconstruction approaches have emerged as a promising alternative means to develop large-area sensors. These methods computationally estimate physical

contact information over the sensor from a few indirect measurements taken by various sensors, as shown in our **HapDef** project (Chapter 2 and Chapter 3).

Electrical resistance tomography (ERT) [151] is a reconstruction technique widely used in large-area tactile sensing to estimate physical contact information using a few spatially distributed electrodes. It typically operates so that a few electrodes are used to inject current into the piezoresistive sensing surface and measure the corresponding electrical potentials (voltages) formed around the current pathway. Physical models derived from Maxwell's equation are conventionally used to reconstruct the resistance distribution, which is usually equivalent to a pressure distribution. However, the physical reconstruction of the pressure map from voltage measurements is an ill-posed and nonlinear inverse problem [152, 153]. The corresponding solution is thus prone to measurement noise and modeling errors. Neural networks can favor achieving a better reconstruction performance [154, 155, 156]. These methods reconstruct the conductivity distribution from the voltage measurements and take the conductivity distribution as the pressure map. This assumption is violated by nature due to nonlinearity in-between.

In real robotic applications, mechanical units, such as force in newtons or pressure in pascals, benefit robot control and mechanical property characterization. Some ERT-based sensors strive to deliver such information instead of pure electrical units, like resistance or conductivity. The studies [157] and [158] obtained remarkable results in predicting single contact locations using machine learning with regression and classification tasks, respectively. However, multiple-contact localization [156, 159] and contact force estimation remain challenging because acquiring sufficient real data requires excessive time and specialized equipment.

We recently explore a real-to-sim transfer approach [7] rather than directly collecting copious experimental data in ERT haptic sensors. It combines a physics-based simulation model with data from a few experiments to learn the mapping from both simulated and real data. In that study, we introduced a multiphysics model of a standard 200 mm $\times$ 200 mm ERT haptic sensor and used it to detect single- and double-contact. It can estimate contact locations with high accuracy but fails in force quantification for double-contact. However, it can not predict more than two simultaneous contacts.

In this chapter, we investigate multiple-contact force estimation using a significantly improved modular pipeline. It involves two stages of mapping: real-to-sim voltage-transfer trained with a matched real and simulated multiple-contact experiments and voltage-to-force inverse mapping trained using the simulated voltages calculated from the force data of the real experiments plus a larger amount of more diverse simulation data synthesized from the multiphysics model.

For this study, we design a new ERT-based sensor with an area of 540 mm by 560 mm. To utilize the proposed approach, we build a testbed with an indenter plate that can collect up to six simultaneous contact data and create a customized multiphysics simulation model to synthesize more diverse contact data. We introduce the sensor design, the testbed, and the simulation model in Section 4.2 and present the results in Section 4.3. We close with a discussion in Section 4.4

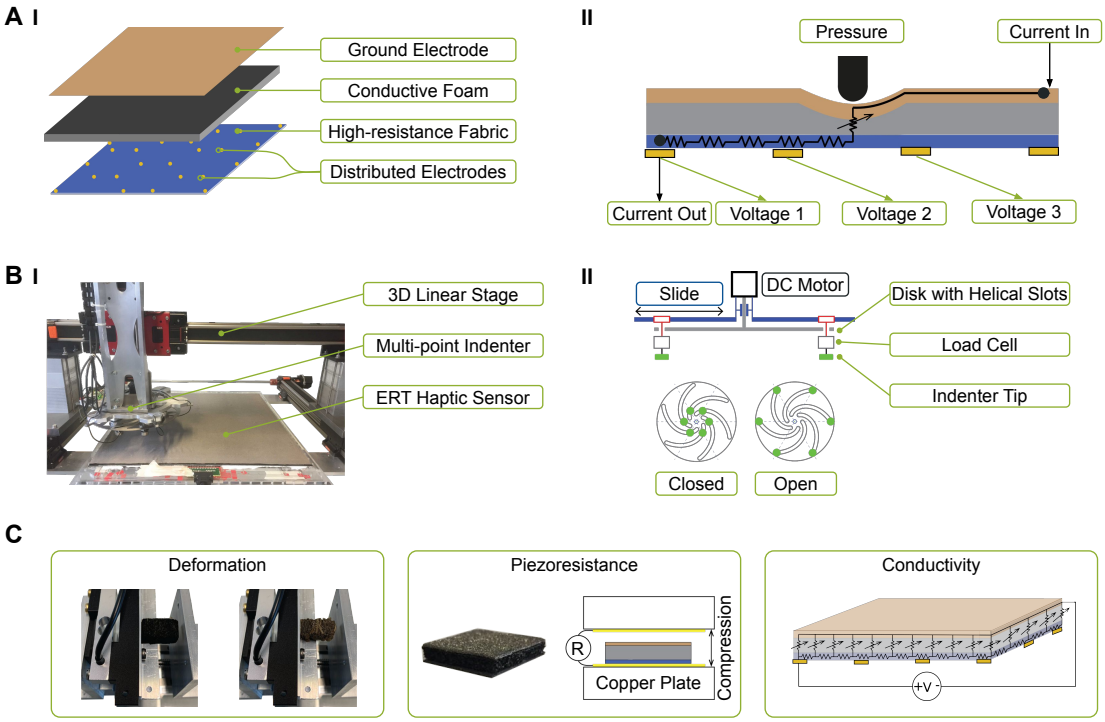

Figure 4.1: **A**: An overview of the newly developed ERT haptic sensor. **B**: Experiment setup (I) and schematics of the multipoint indenter (II). **C**: The multiphysics simulation model comprises three parts: deformation (compression experiments to acquire Young's modulus for both the conductive foam and the top electrode layer), piezoresistance (compression experiment to acquire the piezoresistive behavior of the sensor) and conductivity (schematic illustration).

4.2 Methods

This section describes the hardware setups we built to obtain voltage measurements from a new ERT-based haptic sensor and contact forces from a testbed with a multipoint indenter that probes the sensor, and a multiphysics simulation model that is built for synthesizing more diverse contact data. As the implementation details of the sensor and the simulation model are out of the scope of the author's major contribution to this project, we explain only the essential properties needed to understand the working principle. Implementation details of the design process, circuitry, and electronics are in [7].

4.2.1 Sensor Design

The ERT-based haptic sensor comprises two conductive textile layers and one conductive foam layer, as shown in Fig. 4.1A-I. The top conductive layer is a low resistance fabric (2 $\Omega/\square$) and functions as a large electrode connecting to the ground. The middle layer

is a conductive open-cell foam with high resistance ($10^{11}\Omega/\square$) when undeformed and reduces the local resistance upon being compressed with closed air gaps in the foam. The lower layer is also a high resistance fabric ($10^{4}\Omega/\square$) with 28 sparsely distributed electrodes arranged in a hexagonal pattern. The ERT principle of the sensor works in the following way. A voltage source injects an electrical current from one of the 28 distributed electrodes, and the voltages of all the distributed electrodes are measured, as shown in Fig. 4.1A-II. And the current injection and voltage measurement process repeats until every distributed electrode has been used once for current injection, which results in a vector of 784 voltage measurements (28 distributed electrodes $\times$ 28 voltage measurements per injection).

4.2.2 Testbed

A 3D linear stage with a multipoint indenter is designed as a testbed to conduct indentation experiments and acquire indentation forces along with the corresponding voltage measurements from the sensor, as shown in Fig. 4.1B-I. The linear stage precisely controls the position of the center of the multipoint indenter across the surface of the sensor and in the normal direction to cause contact. The multipoint indenter can hold up to six indenter tips that are each instrumented with a load cell. The load cells can move in the radial direction along linear sliders to vary the relative placement of the contacts, as shown in Fig. 4.1B-II. The radial displacements of the six load cells are controlled together by a single rotary actuator linked to a disk with six helical slots. Each load cell has a cylindrical indenter tip with a diameter of 20 mm. The indenter tips can be manually attached and detached to set the number of simultaneous indentations between one and six contacts.

4.2.3 Simulation Model

A simulation model can generate simulated data that complement the real mechanical and electrical data collected from the physical sensor. We build an electromechanical multiphysics model under the assumption of perfect bonding in our laminated sensor. The behavior of the sensor can be modeled with three consecutive physics models that connect contact pressure to voltage measurements.

A linear elastic finite element deformation model is built to calculate the mechanical deformation of the sensor under indentation loads. The mechanical parameters needed for the deformation model are the Poisson's ratios adopted from literature (0.03 for the conductive foam [160] and 0.25 for the top layer [161]) and Young's modulus measured through compression experiments (0.157 MPa for the conductive foam and 0.095 MPa for the top layer), as shown in Fig. 4.1C-Left. The sensor relies on transduction from mechanical deformations to resistance changes of the sensor. Hence, a compression test is performed to acquire the piezoresistive relation thereof, as shown in Fig. 4.1C-Middle. Moreover, a conductivity model describing the ERT behavior is built to calculate the electrical potentials that occur across the bottom high-resistance fabric when electrical

Table 4.1: Summary of the experiments for data collection, with the number of trials for each contact number ranging from one to six, and the sum. Each experiment generates a list of one to six contact locations, the corresponding list of one to six contact force strengths, and a vector of 784 voltages output by the ERT sensor.

Experiment	Source	Contact Pattern	Single	Double	Triple	Quadruple	Quintuple	Sextuple	Total
Real-Hex	Multiple-Contact Indenter	Hexagonal	20 092	21 313	19 712	14 487	15 944	2 126	93 674
Real-Random	Manual Indenters	Random	0	50	50	50	0	0	150
Sim-Hex	Model	Hexagonal	20 092	21 313	19 712	14 487	15 944	2 126	93 674
Sim-Random	Model	Random	33 275	33 450	33 190	33 047	33 630	33 408	200 000

current is injected from one of the distributed electrodes to the top electrode layer, as shown in Fig. 4.1C-Right.

The simulation accuracy of the developed multiphysics model is evaluated by comparing real voltage measurements and simulated voltage measurements. The built simulation model is accurate, with a median fraction of variance unexplained (FVU) of 10%.

4.2.4 Data

For the study in this chapter, we conduct four experiments to collect data, as listed in Table 4.1.

Real data acquisition using the physical setup We attach one indenter-tip to the multipoint indenter plate and collect real experiment data from 3 000 random configurations, each of which includes a random location for the center of the indenter and a random radius for the spacing of the indenter tips. The indenter first moves to the defined location and then vertically probes the sensor surface with a total displacement of 5 mm in 10 incremental steps (0.5 mm each). In total, we collect 33 000 contact trials for the single-contact case. Then we increase the number of the attached indenter-tips from one to six and repeat the data collection procedure. After collecting the indentation data for multiple contacts, we preprocess the data by deleting the contacts with force strength lower than 5 N concerning the sensor operation range (5 N-40 N). We name this dataset as **Real-Hex**.

Real manual data We additionally prepare three manual indenter types to conduct random multiple-contact experiments with pre-defined loads of 4.905 N, 7.358 N, and 9.810 N, respectively. Each manual indenter has a cylindrical indenter tip with a diameter of 20 mm. We conduct 150 experiments, including 50 double-contact, 50 triple-contact, and 50 quadruple-contact trials. We name this dataset as **Real-Random**.

Simulation for the Real-Hex data using the multiphysics model We provide a simulated counterpart to the Real-Hex data using the built multiphysics model. We compute

Table 4.2: Summary of the datasets, including input and output data that define the associated learning task. The datasets are constructed by combining portions of the experimental data listed in Table 4.1. The variables V_R and V_S stand for voltages from the real and simulated sensors, respectively. The variables M_R and M_S are force maps created from lists of real or simulated contacts.

Dataset Name	Input Variable	Input Experiment(s)	Output Variable	Output Experiment(s)	# of Datapoints
Real	V_R	Real-Hex	M_R	Real-Hex	93 674
R2S-Transfer	V_R	Real-Hex	V_S	Sim-Hex	93 674
S2S-Recon	V_S	Sim-Hex + Sim-Random	M_S	Sim-Hex + Sim-Random	293 674
Real-Unseen	V_R	Real-Random	M_R	Real-Random	150

the simulated sensor's voltages with the same contact locations and force strengths as the Real-Hex trials and name the dataset as **Sim-Hex**.

Simulation data synthesis using the multiphysics model Considering the geometric constraints of the multiple-indenter, we use the simulation model to synthesize widely varying indentation trials. The number of contact points is randomly selected from one to six, with a randomly chosen force range from 5 N to 40 N. In total, 200 000 multiple-contact indentation trials are synthesized. We name this dataset **Sim-Random**.

4.3 Results

We propose a machine-learning framework to infer a force distribution map describing multiple contact locations and force strengths from the raw sensor voltage measurements in a data-driven manner. Conveniently, we can use the data automatically collected in the Real-Hex experiment to directly train an end-to-end model that takes real voltage measurements as its input and provides a force distribution map as its output. We define this machine learning task and the associated dataset *Real* as shown in Table 4.2. However, our indenter design imposes a constrained hexagonal pattern on the contact positions; and it causes all applied contact forces to increase together-these rule out numerous other possible multitouch modes. According to our experience, the end-to-end trained model cannot generalize to other contact modes; we test this hypothesis in Section 4.3.4.

We build the machine-learning pipeline in a modular two-stage format to address aforementioned problem of data insufficiency, as shown in Fig. 4.2. First, we solve the gap between our real experimental setup and our multiphysics model by training a machine-learning model called **Transfer-Net**. It transfers the real voltages (V_R) to match the simulated voltages (V_S) of the multiphysics model under the same contact conditions, essentially learning to eliminate the non-ideal conditions of our physical setup. The dataset we created for training this real-to-sim network is *R2S-Transfer* and contains only voltages from Real-Hex and Sim-Hex (Table 4.2). Learning the inverse mapping, from

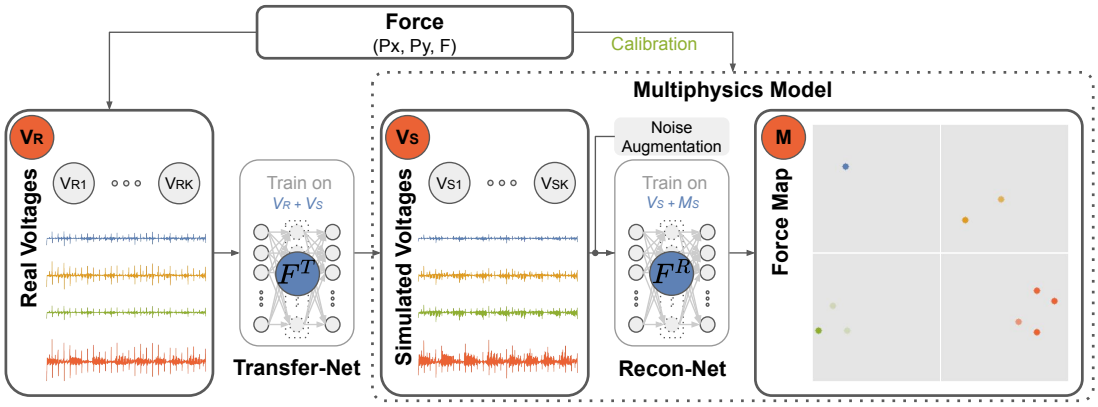

Figure 4.2: We use a two-stage framework to map real voltages to a force map showing where contact is made on the sensor's surface. The first step of this modular pipeline is to transform real data measured from our experimental setup to match the corresponding data generated by our simulation closely. The second step operates on the transformed voltages to output the final force map. Four actual example contact scenarios are shown in blue, yellow, green, and red to illustrate how the 784 real voltages become 784 voltages similar to the simulation and then yield a force distribution map showing the locations and forces of contact.

simulation to reality, is also well-possible. We choose to learn from real to simulation because this division divides the entire mapping into two parts out of two concerns. It turns out to be easier to learn, and it is much easier to generate more simulated data than to obtain more real data.

The second step of our pipeline is called **Recon-Net**, which stands for reconstruction network. It learns to convert a set of 784 simulated voltage measurements into the corresponding force map, which is the output of our entire processing pipeline. Each pixel in the force map contains a value representing the total normal force applied to that sensor region. As listed in Table 4.2, the dataset created for this sim-to-sim learning task is called *S2S-Recon*, which includes the simulated voltages and simulated force maps from Sim-Hex and Sim-Random. To prepare the pipeline for use with real sensor voltages, we perturb the simulated voltages during training with noise mimicking the errors made by the **Transfer-Net**.

The following subsections describe these two steps in detail, evaluate the entire pipeline, and compare our proposed method with two common alternative methods of end-to-end learning and reconstruction based entirely on a physical model.

4.3.1 Transfer Learning

Network structure The Transfer-Net is used to model the differences between our real sensor and the simulation model; this inevitable misalignment is sometimes called the

Table 4.3: Results for the transfer learning task in which the Transfer-Net converts real voltages V_R to the corresponding simulated voltages V_S (*R2S-Transfer* dataset). I, II, III, IV, V, and VI stand for the portions of the datasets with the named number of simultaneous contacts. The gray text indicates extrapolation to an unseen number of contact points, which is expected to perform poorly. The numbers are RMSE / FVU on the training set and the different test sets, with mean RMSE in volts and mean FVU in percentage; lower values of both metrics indicate better inference performance.

Train	I 0.002 / 1.0						I & II 0.002 / 0.5					
Test	I	II	III	IV	V	VI	I	II	III	IV	V	VI
	0.003 / 4.8	0.026 / 75.5	0.031 / 64.4	0.047 / 92.8	0.055 / 97.8	0.047 / 78.7	0.003 / 5.0	0.009 / 4.3	0.019 / 19.4	0.025 / 20.3	0.034 / 34.0	0.029 / 32.1
Train	I, II & III 0.004 / 0.7						I, II, III & IV 0.004 / 0.7					
Test	I	II	III	IV	V	VI	I	II	III	IV	V	VI
	0.004 / 5.1	0.009 / 4.7	0.011 / 4.1	0.019 / 13.2	0.024 / 14.3	0.030 / 16.0	0.004 / 5.1	0.009 / 4.5	0.011 / 4.0	0.010 / 2.3	0.019 / 6.2	0.027 / 12.6
Train	I, II, III, IV & V 0.006 / 0.9						I, II, III, IV, V & VI 0.007 / 1.1					
Test	I	II	III	IV	V	VI	I	II	III	IV	V	VI
	0.004 / 5.2	0.010 / 4.8	0.011 / 4.2	0.010 / 2.4	0.012 / 2.1	0.016 / 5.6	0.004 / 5.7	0.010 / 5.4	0.012 / 5.0	0.010 / 2.5	0.012 / 2.3	0.014 / 4.2

real-to-sim gap. The network has a simple structure of multiple fully connected layers that map the 784 real sensor voltages from a single trial to the 784 simulated voltages for the same contact condition. There are five hidden layers with 784 neurons (ReLU) each.

Data and training details To train the Transfer-Net network, we use the *TransferData* dataset, which provides 93 674 matched pairs of real and simulated vectors of voltages. The dataset is split into training, validation, and test subsets with a ratio of 3:1:1. The Transfer-Net is then trained with the least-squares loss function using the training data set with a batch size of 64 in 1 024 epochs. The loss is optimized using Adam with a momentum of 0.7 and a learning rate of 0.001.

Evaluation We use RMSE and FVU to evaluate the inference performance of the Transfer-Net. ERT-based sensors have a few distributed electrodes and estimate the contacts being applied by measuring the voltage distribution across the piezoresistive sensing surface when current is injected at different locations. Given such sensors' mechanical and electrical design, the system's voltage measurements during multiple contacts cannot simply be estimated by adding together the voltages measured during the individual contacts being experienced. In other words, it is challenging to predict a multiple-contact output from only single-contact data.

We validate this challenge of generalization in ERT-based tactile sensors by first training the Transfer-Net using only single-contact data and testing on all multiple-contact cases. We then incrementally increase the number of contacts used in the training set and compare the generalization performance of the inference. Table 4.3 summarizes these results. As predicted, a network trained on only single-contact data cannot predict the variance in the voltages that occur with two or more contacts. The prediction performance for unseen

Table 4.4: Recon-Net network structure for 160 × 160 output

Layer (type)	Input Shape	Output Shape	# Param
Linear-1	[-1,784]	[-1, 512]	401,920
ConvTranspose2d-3	[-1,512]	[-1, 128, 5, 5]	1,638,528
Upsample-5	[-1,128,5,5]	[-1, 128, 15, 15]	0
ConvTranspose2d-6	[-1,128,15,15]	[-1, 32, 40, 40]	147,488
Upsample-8	[-1,32,40,40]	[-1, 32, 80, 80]	0
ConvTranspose2d-9	[-1,32,80,80]	[-1, 8, 160, 160]	1,032
ConvTranspose2d-11	[-1,8,160,160]	[-1, 2, 160, 160]	146
ConvTranspose2d-13	[-1,1,160,160]	[-1, 1, 160, 160]	19

multiple-contact cases is greatly improved when trials with that number of contacts are included in the training set. The overall performance is high (less than 6% variance unexplained) for all contact numbers when the training set represents the test set.

4.3.2 Reconstruction

Network structure The Recon-Net is designed to map 784 voltage measurements (output of the Transfer-Net) to the corresponding normal force distribution. It is trained on simulated data perturbed with naturalistic noise as described below. Recon-Net has a deconvolutional network structure that maps from the 784 voltage channels to an image-like force distribution map with 160 × 160 pixels. Each pixel in the map corresponds to a 4 mm × 4 mm area of the real sensor describing the distributed force magnitude (in newtons) in the normal direction. The details of the network architecture are provided in Table 4.4. Depending on the needed resolution, the size of the force map can be adapted to 64 × 64 or 32 × 32 with increased pixel sizes of 10 mm and 20 mm, respectively.

Data and training details We train the Recon-Net with the *S2S-Recon* dataset, which contains Sim-Hex (simulated voltages matching the real experiments) and Sim-Random (synthesized force and voltage pairs). As the contact force information is measured only at the center of each of the N contact points, we first need to transform these labels into image-like force maps representing the force applied to each small sensor region. Two parameters define the transformation: the first is the pixel size, which sets the resolution of the force map, and the second is the diameter of the circular area over which each real indentation force is distributed, to account for the fact that the indenter tips are large relative to the default pixel size. The force map dataset is split into training-validation-test subsets with a ratio of 3:1:1 according to contact location, i.e., there is no overlap in locations between the training and test sets. The Recon-Net is trained with a least-squares loss function using the training data set with a batch size of 64 in 1 024 epochs. The loss is optimized using Adam with a momentum of 0.7 and a learning rate of 0.001.

Table 4.5: Recon-Net evaluation on the effects of the two parameters that define the force map: pixel size, which sets the map's resolution, and force distribution diameter, which controls the area to which contact forces are attributed. Number error is the percentage by which the contact number is misestimated; 0% means perfect matching. Location error is the mean distance between each actual contact location and the closest estimated contact. Force error is the mean percentage by which the estimated force deviates from the ground-truth force measurement. Diameter error is the mean percentage difference between the estimated contact diameter and the ground-truth indenter diameter of 20 mm. The mean and standard deviation of the signed values are shown in the table. Negative values indicate underestimation, while positive values indicate overestimation.

Force Map	Map 1	Map 2	Map 3	Map 4	Map 5	Map 6	Map 7	Map 8	Map 9	Map 10	Map 11	Map 12
Pixel [mm] / Diameter [mm]	4 / 20	4 / 28	4 / 32	4 / 44	4 / 52	4 / 60	10 / 10	10 / 30	10 / 50	10 / 70	20 / 20	20 / 60
Number Error [%]	–22±23	–18±22	–14±20	–11±18	–12±18	–12±19	–28±26	–15±20	–11±18	–12±21	–15±22	–4±26
Location Error [mm]	11±22	11±18	11±17	13±16	13±14	15±15	12±25	14±22	13±18	16±19	15±26	23±36
Force Error [%]	–1±52	–7±42	–3±42	5±43	0±40	–1±39	–18±44	–2±42	5±46	–2±41	–7±43	–6±48
Diameter Error [%]	–31±12	11±17	36±18	69±22	107±25	134±28	–34±21	19±17	81±22	171±28	18±15	128±18

Evaluation We evaluate the influence of training data on the performance of the Recon-Net; specifically, we vary the two parameters of the force map (pixel size and force distribution diameter) in twelve different combinations. The resulting Recon-Nets are evaluated using their prediction errors on four criteria: number of contacts, contact location, contact force, and contact diameter. The input to this evaluation is the produced force map. The number of contacts is counted by finding each pixel whose force magnitude is the highest within its circular neighborhood of 64 mm diameter, with a 5 N threshold. After extracting the peaks of the estimated contacts in this way, we use the Hungarian matching method to align the predicted contact locations with the ground-truth locations. The force estimation accuracy is evaluated between the matched pairs by comparing the measured contact force with the sum of all contact forces predicted in the peak's neighborhood. The contact diameter is calculated from the force map based on the force estimates in each contact region using the full width at half maximum (FWHM) criterion, which is similar to RES_{50}, a widely used metric for ERT tactile sensors [152]. This value is compared to the real indenter tip diameter (20 mm).

Table 4.5 shows the performance of the twelve Recon-Nets, and the same results are visualized in Fig. 4.3. Comparing Map 1 to Map 12, where the pixel size and force distribution size are larger, we see the contact number inference improves slightly, while the contact location and diameter inferences perform worse. Comparing from Map 1 to Map 6, the increase in force distribution size tends to improve the force error by reducing the standard deviations. These two parameters thus allow one to tailor the overall system's performance to emphasize either force measurement (contact number and force) or spatial measurement (contact location and diameter).

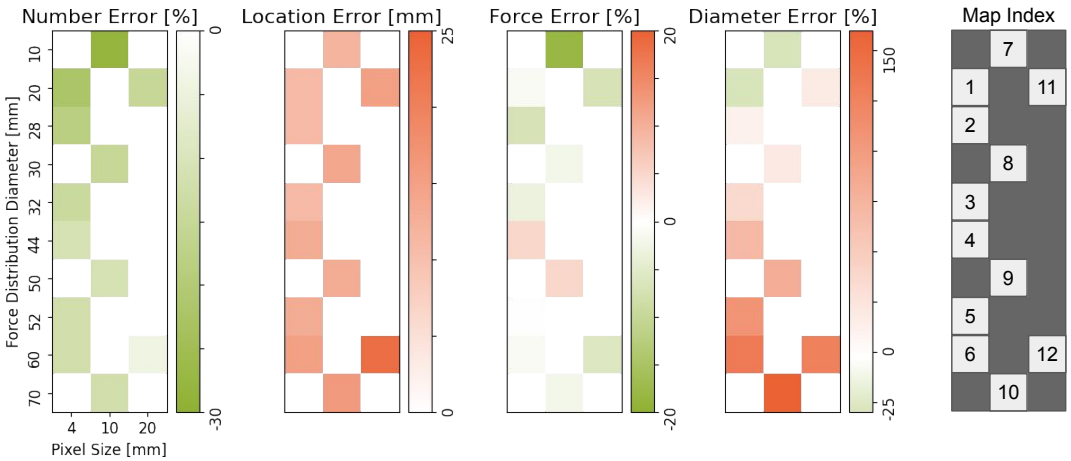

Figure 4.3: Visualization of the mean metrics from Table 4.5 shows how pixel size and force distribution diameter affect the performance of the twelve tested Recon-Nets. The gray areas represent parameter combinations that are not tested.

4.3.3 Complete Pipeline

The whole pipeline integrates the Transfer-Net and the Recon-Net in series. The trained Recon-Net can easily overfit to the simulation data and not generalize well because it is trained only on simulated voltages without considering measurement noise. As reported above, the Transfer-Net is trained on all contact numbers (Table 4.3) and has achieved an average test error of only 4% FVU on previously unseen real voltage data. However, these unpredicted voltage deviations are enough to cause highly incorrect force maps because the reconstruction mapping is highly nonlinear and sensitive to small voltage variations.

Reconstruction robustness To enhance the robustness of the reconstruction process against the real-to-sim voltage-transfer error, we add Gaussian noise to the simulated voltages during training the Recon-Net. We use the residual error of the Transfer-Net to determine the standard deviation of each element in the Gaussian noise vector, as shown in Figure 4.4A. The mean is close to zero for all the 784 measurements, indicating that the training did indeed converge well. Nevertheless, the standard deviation varies from about 0.05 V for well-predicted measurements up to about 0.5 V for a few voltage measurements.

The simulated voltage measurements in the *S2S-Recon* dataset are augmented with different levels of Gaussian noise. Figure 4.4B showcases a triple-contact sample and its force-map predictions trained from three different noise levels. The tested noise levels are 0%, 10%, and 50% of the standard deviations of the transfer error. For this sample, the no-noise scenario predicts a false contact, while the force map of the 10% noise condition is more accurate. The 50% noise prediction is more blurred and dilated relative to the 10%. In summary, the reconstruction tends to be smoother when a higher noise level is used during training. More details are in the following section.

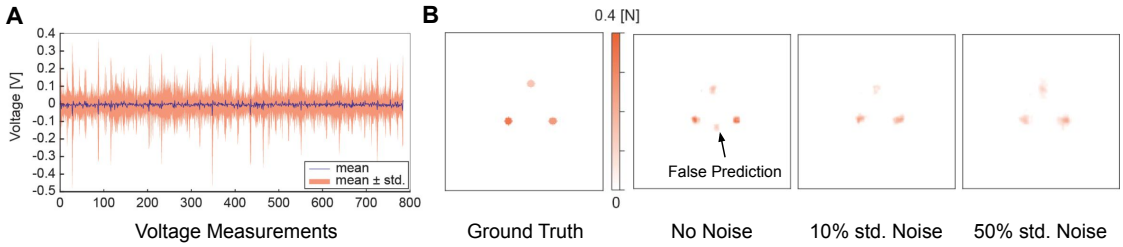

Figure 4.4: **A** shows the mean and standard deviation of the transfer error from the Transfer-Net. **B** presents the force map predictions across the entire surface of the sensor with three different augmented noise levels.

Evaluation We evaluate the whole pipeline with the same criteria used in the ReconNet evaluation. Out of twelve maps (Map 1 to Map 12) with varying pixel sizes and force distribution diameters, four cases are chosen to evaluate the influence of noise augmentation together with the effects of the force-map parameters. The selected ReconNets are Map 1, Map 6, Map 11, and Map 12, which have two different pixel sizes (4 mm and 20 mm) and two different diameter sizes (20 mm and 60 mm). The noise augmentation levels are chosen as 0%, 10%, and 50% of the standard deviations of the transfer error. We use the three-factor analysis of variance (ANOVA) to evaluate the main effects of the three factors and their interactions.

Figure 4.5 presents the force-map inference performance of the whole pipeline. For number error, the main effects of pixel size and force distribution diameter are particularly significant ($p < 0.0001$). Averaged across the three noise levels, the estimated mean of the number error is the smallest for the 20 mm pixel size with 60 mm force distribution diameter (–11.9%); 4 mm pixel size with 20 mm force distribution diameter has the highest average number error (–29.2%). The 10% noise level shows smaller average number error (–14.9%) than the other noise levels (–19.8% and –19.4% for 0% and 50% noise, respectively). The main effects of pixel size and force distribution diameter are also particularly significant ($p < 0.0001$) for the remaining three metrics: location error, force error, and diameter error. The estimated mean of the location error is smaller for the 4 mm pixel size (18.2 mm) than for the 20 mm pixel size (20.3 mm). The 10% noise level shows the largest average location error (20.3 mm compared to 19.0 mm and 18.3 mm for 0% and 50% noise, respectively). In this analysis, the two-way interaction between pixel size and noise level is not significant. The estimated mean of the force error is larger for the 20 mm pixel size with 60 mm force distribution diameter (–13.5%) than the 4 mm pixel size with 60 mm force distribution diameter (–1.9%). The estimated mean of the diameter error is smaller for the 20 mm force distribution diameter (12.1%) than for the 60 mm force distribution diameter (132.3%). The 10% noise level shows the largest diameter error (77.1% compared to 64.2% and 75.6% for 0% and 50% noise levels, respectively).

The noise level shows only a weak effect on the four criteria when the force distribution

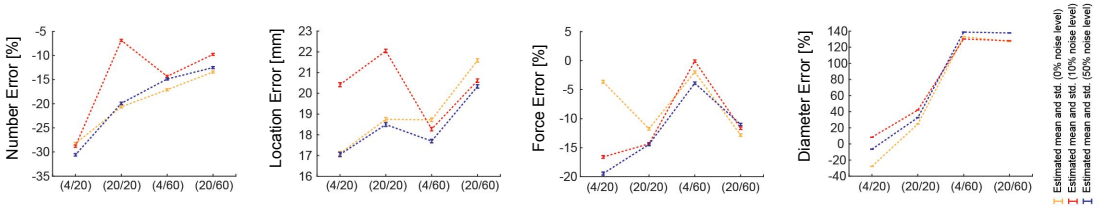

Figure 4.5: Force-map inference performance of the whole pipeline for four Recon-Net configurations (4 mm and 20 mm pixel size, 20 mm and 60 mm force distribution diameter) with three different noise levels (0%, 10% and 50%).

diameter is 60 mm. 10% noise shows better performance in contact number estimation, but worst performance in estimation of contact location and diameter. The estimation of contact force gets worse with the strength of the noise. This performance trend is similar to what happens when the pixel size increases (Table 4.5).

4.3.4 Framework Validation

We use the *Real-Unseen* dataset to evaluate the whole pipeline's performance for unseen contact situations compared to the other two pipelines: direct end-to-end learning from the *Real* dataset and physics-based reconstruction [151]. Figure 4.6A shows two example cases of the force-map predictions obtained by our real-to-sim approach, end-to-end learning and physics-based reconstruction. Our hybrid method predicts the stronger contacts seen in the ground-truth labels. In contrast, although the end-to-end learning method uses the same Recon-Net structure as our method and is trained on the same real data, it results in a complete mismatch between the prediction and the real situation. The physics-based approach tends to show a blurry force map with blobs roughly centered around the ground-truth contacts with very low force strength predictions; its qualitative performance is marginally coherent with the ground truth.

Figure 4.6B compares our hybrid approach with end-to-end learning based on three of our four metrics: location, force, and diameter. The contact number error is not included due to the high location error of the end-to-end approach. This comparison does not show the physics-based reconstruction because its blurry outputs result in erratic evaluation metrics and high errors. The results of our approach and the end-to-end learning are obtained from the twelve different Recon-Net configurations introduced in Section 4.3.2. The results illustrate that the performance of our hybrid pipeline is highly superior to that of both baseline methods. The only metric on which end-to-end learning marginally beats our approach is the prediction error of contact diameter. Among all different variations of our hybrid approach, the best performance is obtained by the Recon-Net with Map 1 and 50% noise augmentation: it achieves average errors of –58.3% in contact number, 74.4 mm in contact location, 17.2% in contact force, and –9.3% in contact diameter.

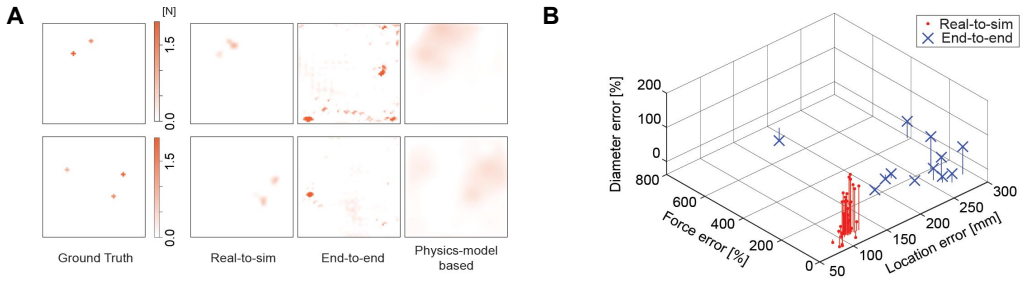

Figure 4.6: **A** shows two example cases of the force-map prediction performance using three different approaches (real-to-sim, end-to-end, and physics-based model). **B** presents the Performance comparison between the real-to-sim obtained from the 36 Recon-Nets (Map 1 to Map 12 with three noise levels) and the end-to-end approaches obtained from the twelve Recon-Nets (Map 1 to Map 12). The physics-based reconstruction is excluded from this comparison due to its inaccurate results.

4.4 Conclusion

Estimating a multiple-contact force map from voltage measurements is a challenging problem in ERT-based tactile sensors whether one uses data-driven approaches or physics-based models. The mapping is nonlinear, and the reconstruction process is ill-posed. We introduce a hybrid approach that learns a data-driven mapping from real voltages of a new soft ERT tactile sensor and simulated voltages synthesized from a multiphysics model. The low-cost multiphysics model simulates voltage measurements caused by tactile stimuli with only a 10% fraction of variance unexplained (FVU), even for multiple-contact cases. The data-driven mapping pipeline comprises two stages: a real-to-sim voltage transfer network trained with a moderate number of real multiple-contact experiments subject to geometric constraints, and a voltage-to-force reconstruction network trained using a larger amount of randomly simulated voltages augmented with noise. For validation, the pipeline is evaluated using a previously unseen dataset containing multiple-contact experiments conducted with the real sensor. When tested with these 150 unseen multiple-contact cases, the proposed pipeline showed an average error of 74.4 mm in contact position (which is about 13% of the sensor width), 17.2% force estimation error, and 9.3% contact diameter prediction error. This result outperforms both a conventional physics-based mapping approach and direct end-to-end learning.

The approach presented in this study can be used to estimate physical stimuli on various ERT tactile sensors and could be adapted to other systems that involve nonlinear inverse mappings. Obtaining the proper inference map of a real large-scale tactile sensor is a difficult technical task. Multiphysics real-to-sim transfer has the potential to greatly ease the development process for many robotic tactile skins by simulating a broad range of tactile inputs that are not subject to the constraints or costs of real experiments.

Chapter 5

Theory and Design of Super-resolution Haptic Skins

The robustness-resolution dilemma limits the wide availability of high-resolution haptic sensors. Previous chapters introduce ways to reduce sensor elements and wiring and to use signal processing (machine learning) to achieve super-resolution accuracy. These designs rely on the empirical knowledge that overlapping multiple taxels' perception fields triggers super-resolution behavior. The design of well-performing sensors is typically achieved by an iterative optimization of all the involved components, namely the mechanical design, data collection, and data processing steps. This process is naturally time-consuming, and it is non-trivial to identify bottlenecks preventing higher performances, especially when optimizing intermediate steps separately. In this chapter, we take on the challenge of proposing a theory for geometric super-resolution to make informed decisions at all levels.

We introduce **taxel-value-isolines**, a novel quantification of a physical sensor response inside a transmission medium. It allows for computing the maximal possible spatial resolution of contact locations and spatial force sensitivity analytically. We show how to approach this theoretical upper bound using machine learning techniques. With our theoretical insights, we design three sensor setups using barometric sensors inside elastomer skin. In correspondence with our analytical predictions, our proof-of-concept sensors achieve a super-resolution factor of 100 in 1-dimension and 1 200 in 2-dimensions, meaning 1 200 times more points than the number of actual measurement elements. Moreover, this chapter provides several guidelines for super-resolution sensor designs, including new materials, structures, and transduction methods. Our theory can guide future haptic sensor designs, inform various design choices, inspire many scientists working on the timely topic of sensor design using AI methods.

5.1 Introduction

Haptic sensing can augment robots' ability to know when, where, and how they interact with other objects [5, 38, 39, 44, 56, 72]. Haptic sensors commonly integrate many small sensing elements (taxels) into a grid, forming either a flat or curved surface, and each element typically senses interactions near its location. Applications for small surfaces

Table 5.1: For the developed prototypes in this chapter we report both the prediction of our theory and the performance of the machine learning inference solution for the 1D case. The performance of the 2D case is also included for comparison.

Sensor Name	Transduction Method	Surface Shape	Evaluation Length/Area	# of Real Taxels	Spacing D [mm]	Data Processing	Output Format	Localization RMSE [mm]	SR Factor[1] Smpl.	SR Factor[1] Avg.	SR Factor[2] Smpl.	SR Factor[2] Avg.
Lepora et al. [127]	Capacitive	2.5D	∼433 [mm^2]	12	4	Bayesian Perception	P	0.12* (MAE)	35	—	∼798	—
TacTip [167]	Cam+Marker	3D	∼2 513 [mm^2]	532	4	Bayesian Perception	P	0.1* (MAE)	40	—	∼150	—
Piacenza et al. [168]	Barometer [104]	2.5D	1 300 [mm^2]	5	∼ 15	Ridge Regression	P	1.6 (MAE)	∼ 9	—	∼32	—
HapDefX	Resistive	3D	24 000 [mm^2]	10	∼ 54	MLP	(P, F)	3.0	∼ 18	—	∼85	—
Piacenza et al. [16]	Optical	3D	∼6 107 [mm^2]	30	∼ 24	MLP	(P, F)	0.6* (MAE)	∼40	32	∼180	123
Hellebrekers et al. [169]	Magnetic	2D	1 600 [mm^2]	5 (3-axis)	15	MLP	(P, F)	0.86 (MAE)	∼17	—	∼138	—
Yan et al. [15]	Magnetic	2D	324 [mm^2]	9 (3-axis)	6	Analytic+MLP+Table	(P, F)	0.1 (MAE)	60	—	∼1146	—
Our Theory in 1D	Barometer	1D	32.5 [mm]	6	6.5	Numeric	(P, F)	0.0525+0.0075	108	448	45	187
Our ML in 1D	Barometer	1D	32.5 [mm]	6	6.5	MLP	(P, F)	0.0416	156	254	65	106
Our ML in 2D	Barometer	2D	676 [mm^2]	25	6.5	MLP$_{(x,y)}$	(P, F)	0.1606	40	52	334	1 260

The sign ∼ indicates our calculation that approximates missing numbers in the literature. RMSE stands for root-mean-square error. MAE stands for mean absolute error. * indicates the best reported performance. [1] uses $\Omega = D/\text{RMSE}$ to calculate the super-resolution factor. [2] uses $\Omega = \text{Length}/(n{\cdot}2{\cdot}\text{RMSE})$ for 1D and $\Omega = \text{Area}/(n{\cdot}\pi{\cdot}\text{RMSE}_x{\cdot}\text{RMSE}_y)$, (Eq. 5.7.2), for 2D to calculate the super-resolution factor. Smpl. denotes the simple form of Ω using the resported RMSE. Avg. stands for averaging the force-specific Ω_k over the entire force range of 0.02 to 1.5 N, where k indicates the k^{th} force intervals (0.01 N).

(e.g., fingertips) and large surfaces (e.g., limbs) sensing need numerous taxels to meet the resolution demand [44, 56, 73, 74, 75, 42, 43]. Fingertips need a high density of fine-sized taxels in a very focused area to access high-resolution information. Large surfaces like limbs require a high amount rather than fine-sized taxels to sense the whole area. Technical challenges arise concerning the physical size of the taxels as well as growing manufacturing and wiring costs [162].

As mentioned and implemented in the above chapters, reducing the taxel numbers while keeping high-resolution performance can be enabled because a sensor element can monitor an extended patch on the continuous surface rather than only a tiny area. We can use particular material properties [127, 129, 9, 76, 15, 126] to spread the contact information to the sensing elements and solve the inverse problem of predicting haptic information from a few taxels, effectively creating high-resolution virtual sensors. This procedure also refers to super-resolution sensing [128].

Super-resolution sensing has been explored in many areas, such as imaging systems (super-resolution microscopy) [125, 163], material releases detection (aerosol/chemical plume release) [164, 165], and geostatistics (temperature and precipitation prediction) [126, 166]. They rely on the continuity and neighboring effects in the transmission medium and solve inverse problems (signal processing) to reconstruct spatial information from a few sensors.

Several haptic sensor designs realize the super-resolution sensing functionality too. A detailed comparison of these sensors is shown in Table 5.1. Lepora et al. [127] developed a Bayesian perception method to localize contact for a capacitive tactile sensor with 12 taxels and achieved a 35-fold super-resolution. A 40-fold SR was achieved by a vision-based tactile sensor (TacTip) by tracking moveable markers with the same method [167]. Piacenza et al. [168] made a dome-shaped sensor by molding five TakkTile [104] pressure sensors under a rubber urethane layer and applied ridge regression to achieve a 9-fold SR. Our HapDefX (Chapter 2 and Chapter 3) used the spreading behavior of mechanical deformation by attaching a few strain gauges on a sizeable robotic limb shell and built a

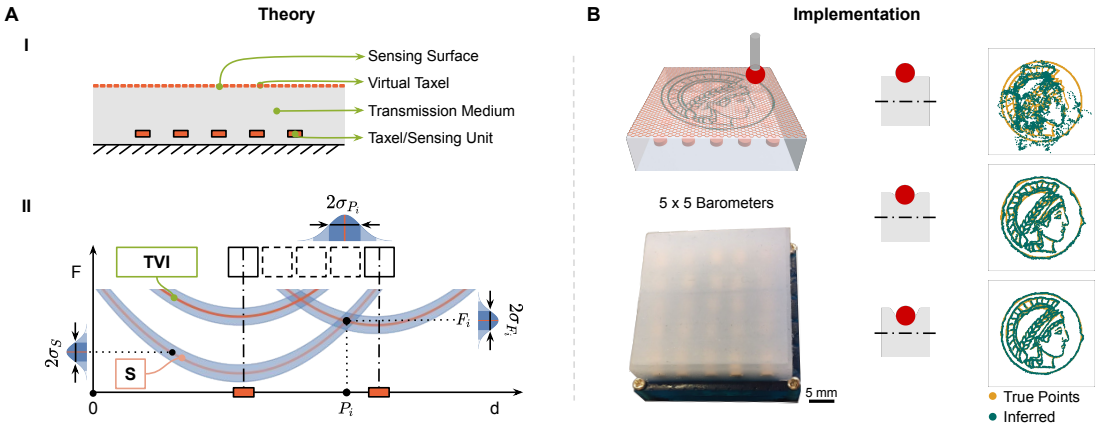

Figure 5.1: Core components of the Theory in **A** and practical implementation in **B**. **A-I** depicts a 1D sensor layout with sparsely placed real taxels and highly resolved virtual taxels. **A-II** shows the orange taxel value isolines (TVIs) that are the force F needed to elicit the same taxel-value S depending on the distance d from the sensor element. The intersection of TVIs of two sensors indicates a hypothetical contact point P_i with a force strength F_i. The measurement noise σ_S in shaded blue leads to uncertainty of the position σ_{P_i} and force σ_{F_i}. More details are given in Fig. 5.2 . **B** shows validation of our theory on a 2D sensor that embeds 25 barometers in an elastomer (EcoFlex 00-30) with a grid layout (5×5) to realize the super-resolution functionality. As a demonstration, a 4 mm red spherical indenter contacts the square sensor surface (34 mm $\times$ 34 mm) following the "Minerva" pattern (diameter 26 mm) with different depths (1.2 mm$\approx$0.2 N, 2.4 mm$\approx$0.6 N, 3.6 mm$\approx$1.0 N). The sensor can resolve the individual contact locations with high accuracy for sufficiently large indentation depths.

machine-learning (ML) model with a structure of multilayer perceptron (MLP) to achieve an 18-fold SR. Piacenza et al. [16] built a soft optic sensor that integrates 30 photodiodes to detect light emitted from 32 paired LEDs. An MLP model was trained to obtain a 40-fold SR. Hellebreker et al. [169] embedded five magnetometers underneath a soft magnetic skin and trained an MLP model to achieve a 17-fold SR. Similarly, Yan et al. [15] used Hall sensors in a sparse configuration to detect the deformation of a magnetized flexible film and developed MLP models to achieve a 60-fold SR.

How much "super-resolution" can we obtain? What is the expected sensitivity of the sensing system? Which material properties are relevant, and which transduction method is advisable? Why and how can machine learning techniques help the inference? This chapter aims to provide answers to these questions by developing a theory that assesses the interplay of the involved physical components, and that is predictive about achievable sensing performance and the therefor required data. Our theory allows us to infer spatially resolved properties, such as accuracy and sensitivity for single contact, based on a single characteristic of the material-taxel interplay: **taxel value isolines (TVIs)**. We demonstrate

the application of the theory by designing two custom-built sensors that yield high super-resolution sensing capabilities. In addition, we provide guidance for three common physical sensor types.

The model we use is a sensor comprising discrete taxels sparsely distributed in a continuous transmission medium underneath a sensing surface, as shown in Fig. 5.1 A-I. For sensing devices that can be described by this model, we derive the single-contact accuracy for position and force magnitude inference based on the taxel value isolines; determine the minimal force profile for which localization is possible; provide conditions for simultaneous contacts to be distinguishable; evaluate the influence of different factors, such as elastic properties of the material, structure design, and transduction methods as well as contacting object shapes; and design a one-dimensional (1D), a two-dimensional (2D), and a three-dimensional (3D) barometer-based sensors to validate our proposed super-resolution theory.

We start with the model description in Section 5.2 and continue with the analysis in 1-dimension in Section 5.3. We then generalize the analysis to 2-dimensions and 3-dimensions in Section 5.4. In Section 5.5, we evaluate different influence factors for the sensor design with help of a finite element method. The isolines of the three real sensor types molded in soft material are presented in Section 5.6. We validate the theoretical analysis with three custom-built sensors and present their performances in Section 5.7.

5.2 The Model

We consider a class of haptic sensing devices intended for measuring force interactions on an extended surface that have an elastic transmission medium covering or embedding physical sensor elements (taxels). A one-dimensional model is shown in Fig. 5.1 A-I. For a point contact, a single taxel value s is a function of the applied contact force strength F, and the displacement d between the contact center and the taxel center:

$$s = f(F,d) + \varepsilon_S, \tag{5.1}$$

where ε_S is the taxel measurement noise with a constant standard deviation σ_S:

$$\varepsilon_S \sim \mathcal{N}(0, \sigma_S^2). \tag{5.2}$$

To relate the sensor measurement noise ε_S to force values F, we assume a linear relationship. Formally, the sensor responses to an applied force at distance 0 should be $f(F,0) = 1/c \cdot F$ and to simplify the notation we set $c = 1$. However, our analysis can be easily adapted for different values of c and for nonlinear monotonic relationships. A detailed analysis of c for a sensor type used in this paper is provided in Section 5.6, which supports our linearity assumption. One of our main contributions is to introduce the *taxel value isolines* as an important characteristic function of the system.

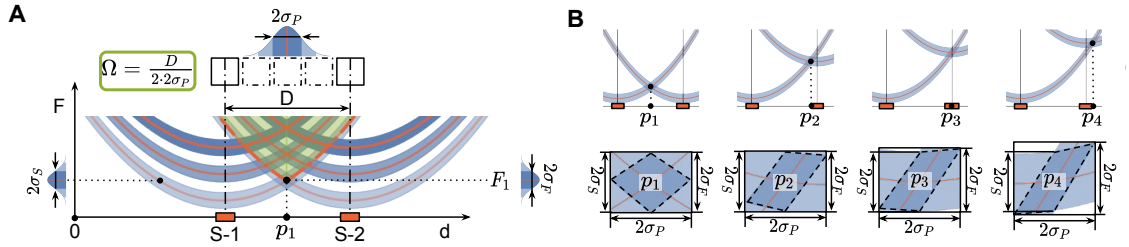

Figure 5.2: Model for contact point localization at super-resolution (Ω). **A** shows the intersection of TVIs of two sensors (marked as S-1 and S-2). The place where the lines corresponding to the particular sensor readings cross is the contact location that is marked as p_1. The measurement noise σ_S leads to uncertainty of position σ_P and force σ_F. **B**: Different intersection types.

Definition 1 Taxel value isolines (TVIs) *are a family of curves*

$$I^S(d) = \begin{cases} F & \text{with } f(F,d) = S \\ \text{undefined} & \text{if no such } F \end{cases} \tag{5.3}$$

where the mean taxel output (Eq. 5.1) has a constant value S.

The TVI for the model system and the effect of measurement noise are shown in Fig. 5.1 A-II. Intuitively, the isolines quantify how much force is needed along the surface to yield the same sensor value. To activate a taxel with a particular value, the required force strength is smaller when the contact location is closer to the taxel. Based on these isolines, we can derive the accuracy distributions of contact localization, force quantification, and the distribution of sensitivity over the sensing surface. Our method is based on the conditions for unique nonlinear triangulation and the geometry of intersecting isolines with their uncertainty bands, as shown in Fig. 5.3.

5.3 Super-resolution in 1D

When can a single contact point be localized at super-resolution, meaning much more accurate than the distance between taxels? Intuitively, this is possible when two or more taxels measure nonzero responses to the contact force and the activation pattern of the taxels is unique for this location. More formally, this condition can be analyzed with the TVIs. The minimal setting of two taxels located at a distance of D is shown in Fig. 5.2A.

A particular contact event causes a sensor reading in both sensors S-1 and S-2. The TVI corresponding to a sensor reading of S-1 relates the position to the force of the potential contact point. Only if the TVIs from both sensors intersect, the contact position can be localized, up to some uncertainty introduced by the measurement noise σ_S (Eq. 5.1). With this insight, we can derive the minimal force F_S (force sensitivity) that allows for super-resolution localization, i. e., where at least two isolines intersect. We can find this minimal force, where the isoline corresponding to the smallest taxel sensitivity S_{min}

intersects with TVIs from neighboring taxels. The area where $F \geq F_S$ is green shaded in Fig. 5.2 A.

As we assume an additive measurement noise, the uncertainty is also added to the isolines. What does this mean for the ability to localize the contact position and to infer force magnitude? For the case of parabolic TVIs, we have to consider different possible scenarios, as shown in Fig. 5.2 B. For contact locations between the two taxels, marked with p_1, p_2, the standard deviation for the force estimation is actually constant and given by σ_S. There is no dependency on the taxel distance D. For contact points close to the taxel and outside the taxel pair, the uncertainty of the force magnitude grows with distance (p_3, p_4). The standard deviation for position localization σ_P is typically largest in the center between the taxels and gets smaller on either side (details and exceptions are in the following).

Knowing the position accuracy σ_P, we can quantify the super-resolution capabilities. For simplicity, we use $2\sigma_P$ as the size of a virtual taxel, as shown in Fig. 5.2 A, corresponding to a confidence interval of about 68%. Thus, the spatial resolution is $2\sigma_P$. Between two sensors at a distance D we can distinguish $D/2\sigma_P$ virtual taxels, which we define as *super-resolution factor* w.r.t. real taxel number n ($n = 2$ in Fig. 5.2 A):

$$\Omega = \frac{D}{n \cdot 2\sigma_P}. \tag{5.4}$$

As summarized in Table 5.1, another commonly used metric for the SR factor divides D by the localization prediction error (root-mean-square error (RMSE) or mean-absolute error (MAE)).

5.3.1 Influence of the Isoline Shape

The shape of the TVIs depends on the properties of the transmission medium and the sensor element type. We study the impact of the TVI shape on the accuracy of single contact force inference. We follow the general model that the response of a taxel to a force on the surface decreases monotonically with distance from the taxel center. This in turn leads to a monotonically increasing TVI.

We assume the attenuation of taxel response behaves as $s \propto 1/|d|^\alpha$ for a force at distance $|d|$ from the taxel center with the attenuation exponent α. Without loss of generality, we consider two taxels at a distance of 1. Their TVIs are given by

$$I_1^{S_1}(d) = g(S_1) + |d|^\alpha, \tag{5.5}$$

$$I_2^{S_2}(d,s) = g(S_2) + |1-d|^\alpha. \tag{5.6}$$

where $g(S)$ is the force corresponding to the measurement at distance 0. The TVIs for different attenuation behaviors are shown in Fig. 5.3 A. The accuracy of super-resolution sensing is strongly affected by α. For instance, for linear attenuation ($\alpha = 1$), the local-

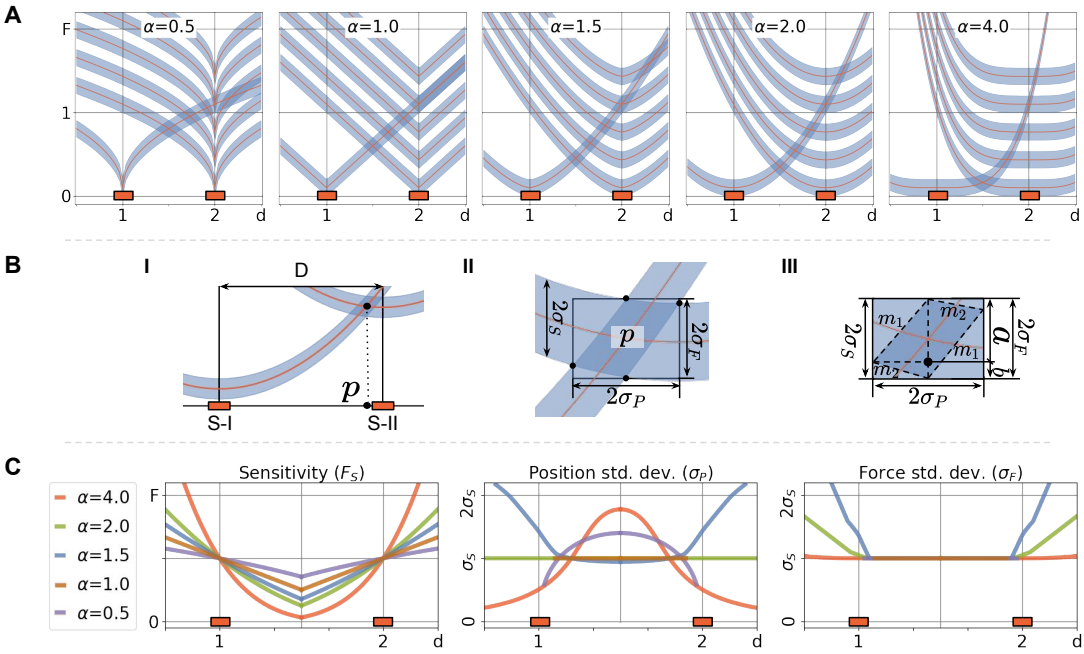

Figure 5.3: Effect of TVIs shape on super-resolution characteristics. **A**: TVIs for two taxels (at distance 1) with different attenuation exponents α. Note the different shapes of the intersection areas. **B**: Computation of σ_P and σ_F. **B-I** shows the intersection of two TVIs caused by a single contact at position p. **B-II**: Precise computation of σ_P and σ_F based on the intersection type. **B-III**: Approximated computation of σ_P. The intersection area of two TVIs between the taxels can be well approximated by a parallelogram. The slopes of the edges (m_1, m_2) are simply given by the inclination of the TVIs at the intersection point (first order approximation). **C**: Resulting sensitivity, standard deviation of position localization, and standard deviation of force inference with different attenuation exponents α.

ization will only be possible between the taxels, but not outside (unbounded intersection area). For concave curves ($\alpha < 1$), there can be three intersection areas: one is between the two taxels and the other two are outside, which makes the reconstruction not unique. For the sake of comparison, we assume that in this case the correct intersection is known.

Precise computation of σ_P and σ_F For two taxels at distance D, as shown in Fig. 5.3 B-I, we use the exact corners of the intersection area of the two TVIs with their measurement uncertainty hose. These intersection points are computed numerically, and we use this method for all plots in this chapter. In more detail, for each TVI we have an upper and lower bound for the uncertainty (one standard deviation) denoted by $h^{\pm}$:

$$h_1^+ = I_1^{S_1}(d) + \sigma_S \tag{5.7}$$

$$h_1^- = I_1^{S_1}(d) - \sigma_S \tag{5.8}$$
$$h_2^+ = I_2^{S_2}(D-d) + \sigma_S \tag{5.9}$$
$$h_2^- = I_2^{S_2}(D-d) - \sigma_S. \tag{5.10}$$

There are four intersection points: $h_1^+ \cap h_2^+$, $h_1^+ \cap h_2^-$, $h_1^- \cap h_2^+$, and $h_1^- \cap h_2^-$, which we illustrate in Fig. 5.3 B-II. The bounding box size of these four points defines σ_P and σ_S.

Smaller attenuation exponent α yields a smaller sensitivity (bigger F_S) between the taxels but in general a more homogeneous distribution (Fig. 5.3 C). The position accuracy, measured by the standard deviation, σ_P is constant for $\alpha = 2$. For normalized TVIs, as used here for comparison, $\sigma_P = \sigma_S$. The general case is given below. For $\alpha > 2$ and $\alpha < 1$ we find a reduced position accuracy between the taxels. Interestingly, for $\alpha > 2$, the reduced position error at the outside is paid by an increase in minimal force (reduced force sensitivity). The ability to infer the force magnitude, given by σ_F, is constant between the two taxels, irrespective of α. For contact points outside, larger values of α are better, whereas for $\alpha \leq 1$ no detection is possible. For $1 < \alpha < 2$, the position error and force error are both high outside. From this analysis, $\alpha = 2$ yields the best overall characteristics, namely having high force sensitivity (small F_S), high position, and force accuracy (small σ_P, σ_F respectively). Values $\alpha > 2$ trade off some super-resolution by sensitivity, but are generally also good. Taxel response attenuation with exponent $\alpha < 2$ should be avoided, if possible.

Approximate analytical computation of σ_P Here, we give a closed-form expression relating the force measurement noise to the position accuracy between two taxels. We make a first-order approximation of the intersection area as a parallelogram, as shown in Fig. 5.3 B-III. For two taxels at distance D and the TVIs described by Eq. 5.5 and Eq. 5.6, we can compute the (absolute) derivatives of the TVIs at the intersection point d:

$$m_1 := \left| \frac{\mathrm{d}I_1^S(d)}{\mathrm{d}d} \right| = \lambda \alpha |d|^{\alpha-1} \tag{5.11}$$
$$m_2 := \left| \frac{\mathrm{d}I_2^S(D-d)}{\mathrm{d}d} \right| = \lambda \alpha |D-d|^{\alpha-1}. \tag{5.12}$$

The trigonometric construction is shown in Fig. 5.3 B-III and yields the following result:

$$2\sigma_S = a + b = m_1\sigma_P + m_2\sigma_P \tag{5.13}$$
$$\sigma_P = \frac{2\sigma_S}{m_1 + m_2}. \tag{5.14}$$

For simplicity, we can use position accuracy at $d = D/2$ (which is the worst for typical exponents, $\alpha \geq 2$) for $n = 2$ sensors to quickly compute the super-resolution factor

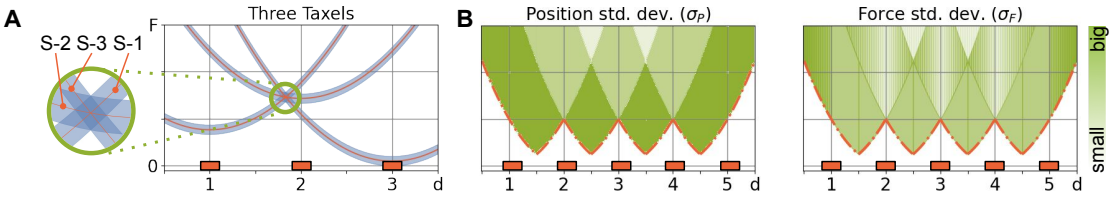

Figure 5.4: Theoretical super-resolution characteristics of a 1D sensor with multiple taxels. **A**: Three taxels localizing a single contact. **B**: The spatial distribution of accuracy for a single contact. Below the orange dash-dotted line, no super-resolution localization is possible. Notice the increase in accuracy for higher forces because multiple taxels are activated.

analytically using Eq. 5.4:

$$\Omega = \frac{D\lambda\alpha(D/2)^{\alpha-1}}{2\cdot 2\sigma_S} \tag{5.15}$$

To be more accurate we need to consider all positions between taxels instead of the middle one, as done in Fig. 5.8 D-I and detailed in Section 5.7.1.

5.3.2 Multiple Taxels in a Line

A real sensor should clearly contain more than 2 taxels, so let us consider multiple taxels (with quadratic isolines) equidistantly placed in the transmission medium (as shown in Fig. 5.1 A-I). As expected, in the area where more than two taxels respond to the contact stimulation, higher localization accuracy is possible, as shown in Fig. 5.4 A. This leads to a reduction in uncertainty about the contact force due to two effects. The first one is the averaging of independent noise sources leading to a factor of $\sqrt{1/n}$, where n is the number of participating taxels. The second effect comes from the intersection of TVIs from more distant taxels, which, due to a higher TVI slop, results in a lower uncertainty as apparent from Eq. 5.14. The sensitivity F_S is shown in Fig. 5.2 B as a dash-dotted lower bound, together with the accuracy of localization σ_P and force quantification σ_F. The sensitivity is not homogeneous and is higher between taxels, which might be a surprising result at first glance. In summary, the most important take-home message is to have multiple taxels responding to a contact force because it improves accuracy. Closer placement of taxels increases both sensitivity and accuracy, whereas reducing taxel measurement noise improves accuracy but not sensitivity.

5.3.3 Multiple Contact Points

In many applications, we are interested in detecting multiple simultaneous contact points. In this section, we investigate the criterion to discriminate double contacts. At first glance,

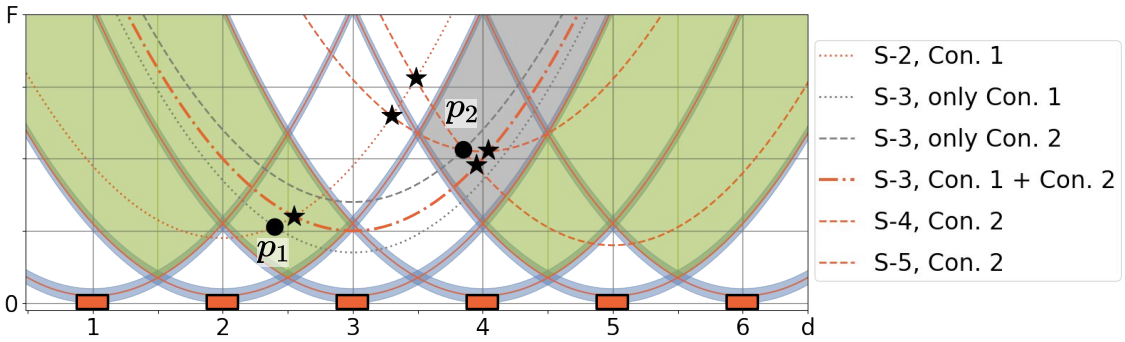

Figure 5.5: Multiple contacts discrimination. There are six sensing elements aligning in one line forming a 1D layout to discriminate double contacts. When one contact locates in the left green-shaded area, all contact situations of the second contact can be successfully discriminated when they are in the right green-shaded area. In contrast, when the second point is too close, say in the gray-shaded area, the following problem occurs. The first contact at p_1 alone would activate two sensing elements (S-2: orange-dotted line, S-3: gray-dotted line). Another contact in the gray-shaded area at p_2 would activate three sensing elements (S-3: gray-dashed line, S-4: orange-dashed line, S-5: orange-dashed line). When these two contacts happening simultaneously, the value of the sensing element S-3 turns to be an orange-dash-dotted line (combined TVI). This will result in an unreliable inference. Multiple spurious contact points occur ("★") rather than two actual contact points "•")

we need 2 pairs of taxels. However, when the contact points are too close, spurious contact points would be detected due to additional intersections of TVIs. The basic condition for successfully distinguishing them is shown in Fig. 5.5. We consider two contact points with individual force strength are located in a 1D sensor layout with six sensing elements. When one contact is located in the left green-shaded area, then a second contact can be clearly discriminated when it is within the right green-shaded area. In contrast, when the second contact is located in the gray-shaded area, it activates at least three sensing elements (S-3,S-4,S-5). Notice, the sensing element S-3 is also activated by the first contact. Thus, the value of S-3 is then not reliable for triangulation inference and will create spurious inference points and modify the position of the otherwise correctly inferred locations. The most important take-home message is to have at least two taxels between these two contacts and at most one taxel of them that is evoked by one contact not the other. With higher forces applied by these two contacts, the distinguishable distance for double simultaneous contact is larger.

5.4 Super-resolution in 2D

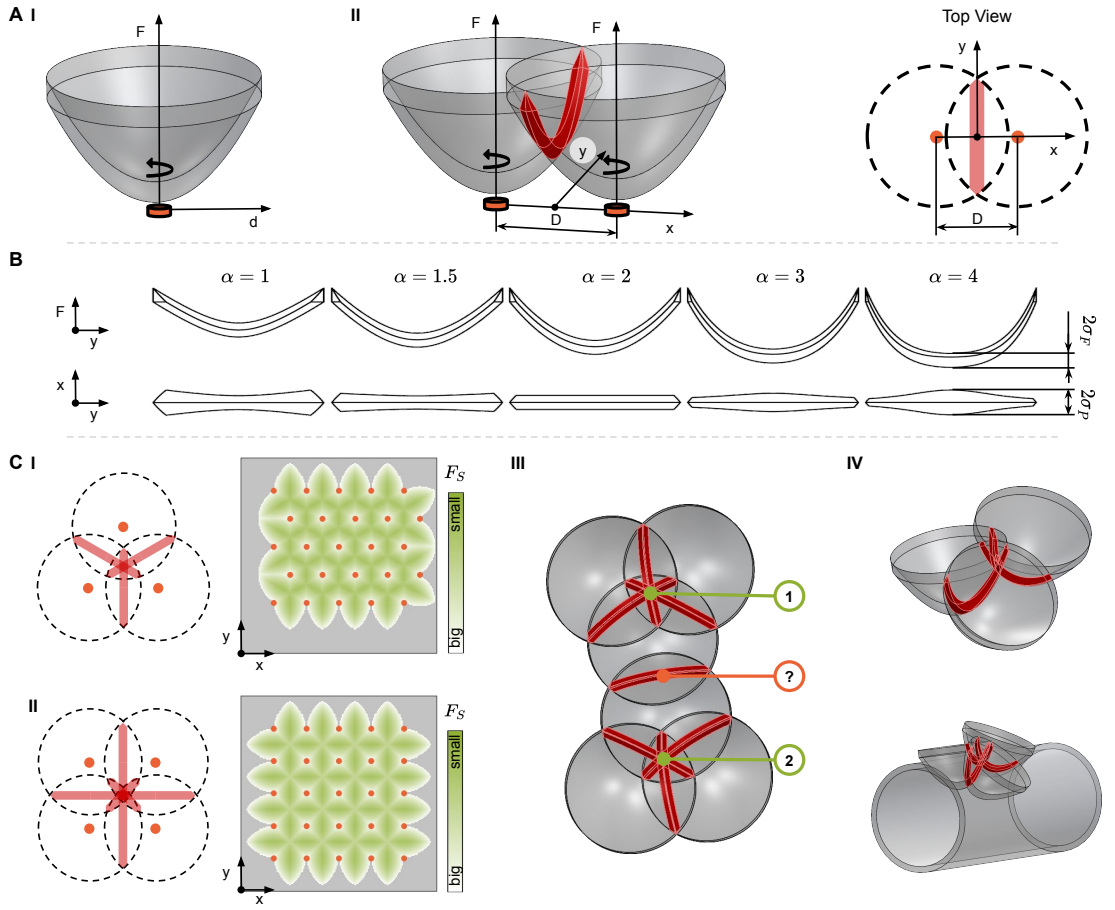

Figure 5.6: Taxel values isolines for a 2D sensing surface. **A-I**: Taxel value isoline (iso-surface) for a single taxel. **A-II & III**: Intersection volume (due to measurement uncertainty) for two taxels at a distance D along the x axis, see top view. The localization would be very uncertain along the y direction. **B**: Intersection volume for different attenuation exponents α. **C**: Proper localization requires at least 3 taxels for 1 contact point. **I** shows a hexagonal sensor placement with resulting sensitivity over the surface and an illustration of the intersection volume of the TVIs from 3 taxels. **II** is the same for a grid and 4 taxels. **III** illustrates a spurious contact localization for 2 contact points and 6 taxels (marked with "?"), however its uncertainty would be very large. **IV** illustrates the intersection of TVIs for a curved sensing surface. The same applies as in the planar 2D case, except that distances need to measured as geodesics on the curved surface.

The above analysis for the 1-dimensional case helps us to investigate a sensor with a flat or curved 2-dimensional sensing surface. To simplify the analysis, we continue to assume a homogeneous transmission medium and an isotropic sensor element, which is, for instance, approximately true for barometric sensors but violated for strain-gauge sensors. We first consider a flat 2D sensing surface with coordinates x and y. The concept of isolines translates into isosurfaces, as shown in Fig. 5.6 A-I. However, we still call them TVIs for consistency.

Clearly, with only two taxels, an accurate localization can only be done along one dimension, but not along two. The intersection volume of the TVIs is presented in Fig. 5.6 A-II & III. For attenuation exponents, $\alpha \neq 2$, the intersection volume changes depending on the transversal position y, as shown in Fig. 5.6 B.

To make proper super-resolution localization possible, at least 3 taxels need to respond to a stimulus, as shown in Fig. 5.6 C-I. As before, the accuracy increases if more taxels are involved, as shown in Fig. 5.6 C-II for four taxels arranged in a regular grid. We also compute the sensitivity distribution over the sensing surface for both sensor arrangements. The honeycomb pattern allows for a more homogeneous sensitivity.

To detect two simultaneous contacts, at least 6 taxels are required to respond. Similar to the 1D case, if the contact points are too close, spurious intersections can occur. In some cases, as shown in Fig. 5.6 C-III, spurious intersections can be ruled out because of high elongation in one direction. This is a new feature that was not observed in 1D. Very similar considerations are also valid for curved sensing surfaces, which is illustrated in Fig. 5.6C-IV.

5.5 Physical Factors Influencing TVIs

We investigate the detailed influences of the properties of the transition medium (e.g., elastomer) and the taxel placement on the TVIs and on the sensitivity. External contact at the sensing surface causes deformation of the transmission medium that can be measured by physical sensor elements (taxels). In this section, we assume the taxels are able to measure deformation/strain within the medium, either isotropically or directionally.

In principle, the deformation can be described by the absolute movement of elements, called displacement, and by the relative movement of elements, denoted as strain. To study the isolines we consider the displacement as a measure of deformation. The displacement is computed for a static mechanical model simulated with the finite element method (FEM) using Ansys [136]. The questions we want to answer in this section are: Where to place the sensor elements within the transition medium? How thick should be the transition medium? What is the effect of material properties such as the Poisson's ratio?

Simulation model The full model is a cylinder-shaped sensor transmission medium with a diameter of 200 mm and a thickness of 5 mm, 10 mm, or 15 mm, but we simplify the model in the FEM simulation into a 2D plane due to the axis-symmetric property of

the cylinder and the spherical indenter (10 mm diameter) as shown in Fig. 5.7 A-I. The transmission medium is constrained at the bottom by a fixed boundary. We analyze the displacement when a normal force is applied to the contact surface via the indenter. In Fig. 5.7 A-II the displacement is evaluated in x- and z-direction, as well as in terms of total displacement. Considering the displacement at three different depths, close to the sensing surface (L1), in the middle (L2), and close to the fixed boundary (L3), shows significant differences. The displacement in depth direction (Δz) decreases monotonically with distance, but is much stronger at L1 than at L3. The radial displacement (Δx) first increases and then attenuates, which will give rise to non-monotonic TVIs.

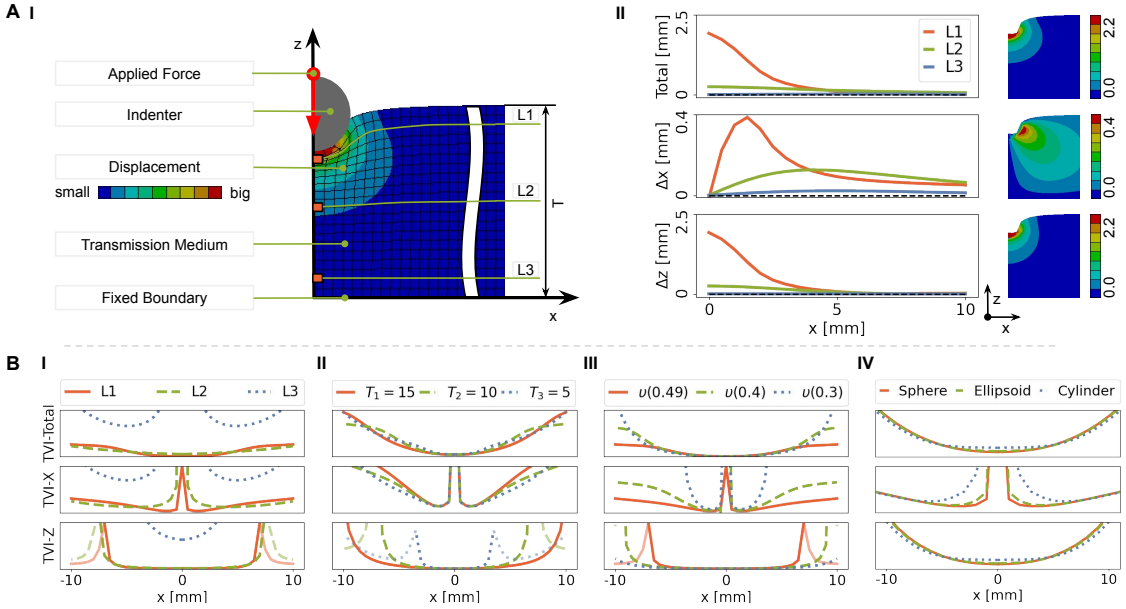

Figure 5.7: Physical Factors Influencing TVIs using FEM. **A-I**: The model: because of symmetry, only one half is shown. The thickness is $T = 10$ mm, and we consider taxels at level $L1 = T - 1$ mm, $L2 = T - 5$ mm and $L3 = T - 9$ mm. **A-II**: Total displacement maps as well as its x and z components. **B**: TVIs for hypothetical sensors measuring total displacement or component-wise displacement, for different taxel depth (**I**), material thickness (**II**), Poisson's ratios (**III**), and indenter shapes (**IV**). TVIs for negative sensor values are in a light shade. Default configurations are: Poisson's ratio $\nu = 0.49$, Young's modulus $E = 0.07$ MPa, and density $\rho = 1.07$ g/cm^3 for the transmission medium; Poisson's ratio $\nu = 0.33$, Young's modulus $E = 71$ GPa, and density $\rho = 2.77$ g/cm^3 for the indenter; bonded contact type without friction consideration.

Taxel depth The dependency of the deformation on the depth has an impact on the TVIs. Figure. 5.7 B-I shows the TVIs for ideal displacement sensors at three different depths (L1-L3) described above. We plot the TVI for the smallest non-zero taxel value.

However, because there can also be negative displacements, we add a light-shaded TVI for small negative values if occurring. When measuring depth deformation, the TVIs have a convex shape for positive sensor values and a second parabola at either side for the negative sensor values. This holds for all depths, although only visible for L1 due to a limited plot range. The radial direction is more surprising, as the center is not moving radially, the TVIs have a pole in the center. The total deformation is monotonic for low to middle taxel depths (L1 and L2). Depending on the physical quantity measured by the real sensor element, this leads to different conclusions: For strain gauges and accelerometers, for instance, the global displacement can be considered as a good proxy, as they measure the curvature and inclination of the material. Thus, these sensors should be more close to the sensing surface. Barometric sensors, for instance, measure local displacement, for which we cannot give a recommendation at this point.

Sensing surface thickness Guided by the previous result, we consider a taxel at depth L1 (1 mm below the sensing surface) for different total thicknesses of the elastomer. Figure. 5.7 B-II presents the TVIs. The radial deformation is, as expected, only a little effected by the amount of material underneath the taxel. The depth deformation, on the other hand, shows a much larger effect of thickness, because for smaller thickness we get less displacement. In summary, the thicker material causes more displacement and in turn increases the sensitivity, because the fixed boundary becomes less dominant. Notice, for thicker material, the shear/x-directional displacement has stronger attenuation property and has less shear sensitivity. Nevertheless, there are several ways to increase the sensitivity in the shear direction, as proposed in earlier work [170, 10]. One is to add ridges on the contact surface [170]. Another is to use multiple layers of materials to cause mechanical contrast and affects the force distribution [10].

Material properties For a robust sensor design, we need to make sure that the contact-caused stress is smaller than the yield strength of the transmission medium, and the material deformation is an elastic behavior. The main two properties of the transmission material we consider are the Young's modulus and the Poisson's ratio. The Young's modulus describes how easy the material deforms and has a proportional impact on the deformation. A soft material (small Young's modulus) improves sensitivity, but is also increasingly deformed by inertial effects. Poisson's ratio measures the relative transversal/radial expansion when the material is axially compressed. Most elastomers have Poisson's ratio around 0.5 and metals have around 0.3 [171]. Figure. 5.7 B-III shows the TVIs for different Poisson's ratios. Decreasing the ratio, the radial displacement becomes much less sensitive, whereas the depth displacement has higher sensitivity (lower TVIs). Thus, depending on the measurement direction of the real sensor element, different Poisson's ratios are preferred.

Indenter shape What is the impact of the indenter shape (or the object getting in touch with the sensing device later) on the TVIs? For illustration, we simulate the displacement caused by three different indenter shapes: a sphere, an ellipsoid, and a cylinder. The displacement decreases and becomes flatter for the cylinder near the contact center, as shown in Fig. 5.7 B-IV. The cylinder with flatter contact surface causes less sensitivity for shear displacement. Increasing the indenter size, the displacement becomes smaller, which results in less sensitivity too. As a remark, for too small indenters, there is a risk to exceed the material yield strength and break the contact surface material.

5.6 Taxel Value Isolines of Real Sensors

We complement our theoretical analysis by measuring the response curves of three suitable sensors. We consider strain gauges that measure the change in curvature along one direction averaged over the sensing area [129, 9]; accelerometers that are able to measure the absolute inclination of the local elastomer patch using the gravity direction as a reference [172]; and barometers that sense the volume change caused by the material's deformation in the form of isotropic pressure [104, 106].

Experimental setup As shown in Fig. 5.8 A, we mold a single sensor element at the center of an elastomer disc (120 mm diameter and 10 mm thickness) and measure the sensor element's response when indenting it with a controlled depth along two perpendicular axes using an automated 3-DoF testbed with a force sensor. The strain gauge sensor is EP08-250AE-350, the accelerometer is ADXL345, and the barometer is MPL3115A2. We solder extra thin wires (CU-enameled wire with a diameter of 0.15 mm, ME-MeßSysteme GmbH) to the chips to be able to mold them inside the elastomer with minimal mechanical influence of the taxel. We mold these three sensors floating in the middle and center of three elastomer pieces (EcoFlex 00-30) individually. Additionally, we build a 3-DoF testbed (Barch Motion, resolution: 0.0075 mm) equipped with a 3-axis force measurement device (ATI Mini40, resolution: $0.01/0.01/0.02\,\mathrm{N}$ ($F_x/F_y/F_z$)) to indent the elastomer surface in a precise and automated way.

Data collection As shown in Fig. 5.8 A, a 4 mm spherical indenter goes along two black-colored perpendicular trajectories where each sensor element locates under the cross point. The testbed is used to make the indenter contact 101 positions evenly spread along each trajectory (from -50 mm to 50 mm) with 25 incremental indentation depths (0.2 mm each) at each position. The sensor value of the strain gauge (EP08-250AE-350) is acquired through a quarter Wheatstone bridge and amplified by an MCP609; the sensor values of the accelerometer (ADXL345) and the barometer (MPL3115A2) are acquired through the evaluation boards supplied by Adafruit; and all of them are delivered to a personal computer through an Arduino Mega 2560. The recorded data are sensor values, force values, and indentation positions and depths. For different indentation depths, the

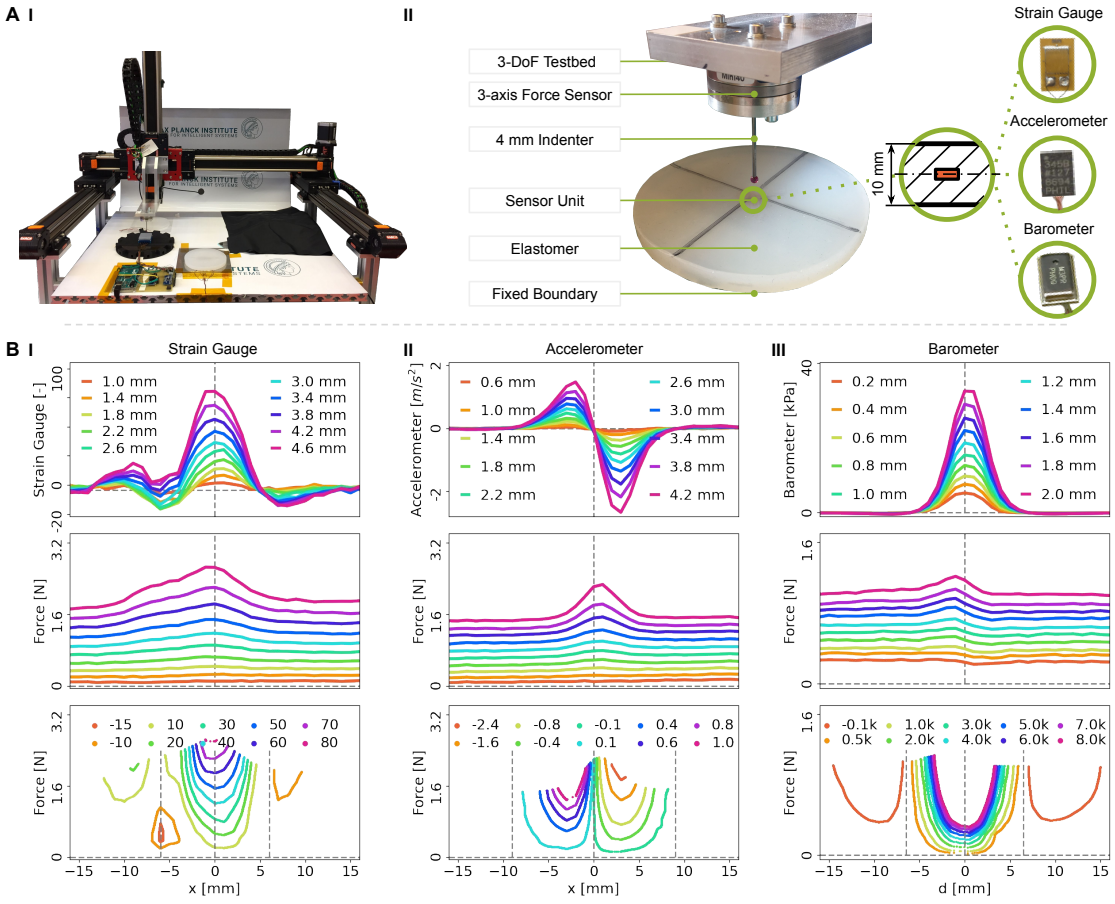

Figure 5.8: Response and TVIs of real sensor elements. **A**: The experimental setup with the whole testbed in **I** and elastomer with a single taxel in **II**. The black lines on the disc mark the stimulation points. **B**: Sensor response (top), true indentation force (middle), and TVIs (bottom) for strain gauge (left), accelerometer (center), and barometer (right). Lines are presented for different penetration depths of the indenter and different sensor values in the bottom row. The barometer has TVIs that are most similar to our theoretical model.

sensor values vary along the indentation positions, and deeper indentations need higher forces and result in higher sensor values. The applied forces are higher at indentation positions near the placed sensor elements because the physical sensor element is stiffer than the elastomer.

Data processing Based on the recorded data, we present in Fig. 5.8 B the sensor values, the applied forces, and the TVIs for each sensor type as a spatially changing quantity. We only show the data for one of the two directions. For the strain gauge element, this is along its measurement direction. Accelerometers are interesting, because they can distinguish between both directions. The barometer is fairly isotropic up to small deviations because of the rectangular sensor housing. We implement the following steps to compute the isolines: first, we linearly interpolate the sensor values and force values; second, we choose a position-related sensor value and find the corresponding force measurement in that position; third, we draw the position-force curve for that sensor value with the same color. The strain gauge has a non-monotonic behavior, where one strain gauge value has several position–force possibilities. The accelerometer does not have this problem, however, it has a "blind spot" directly above the sensor element (no inclination), however, this is a tiny area. Note that the TVIs on both sides are for different sensor values, so super-resolution localization is well possible. We leave a theoretical analysis of these TVI shapes for future work. The barometer shows the convex and monotonic properties as described in our proposed theory. Notice the two orange "wings" which are the TVI for negative values caused by the lateral pulling force from distant contacts.

5.7 Experimental Validation

5.7.1 Quantitative Analysis in 1D

Due to the convexity and symmetry of the barometer's isolines, we use it to validate our proposed theory. We mold six barometers in the elastomer (Diameter: 120 mm, Thickness: 10 mm) along a straight line with approximately 6.5 mm distance to each other as shown in Fig. 5.9 A. The testbed carries a 4 mm spherical indenter, and contacts the surface along the sensor placement center line. Sensor values, force values, and indentation positions and depths are recorded at 5001 positions evenly along the line (50 mm in total, 0.01 mm sample interval) with 40 incremental indentation depths (0.1 mm each) at each position. The resulting taxel responses and the TVIs are presented in Fig. 5.9B. They suggest a very good super-resolution potential.

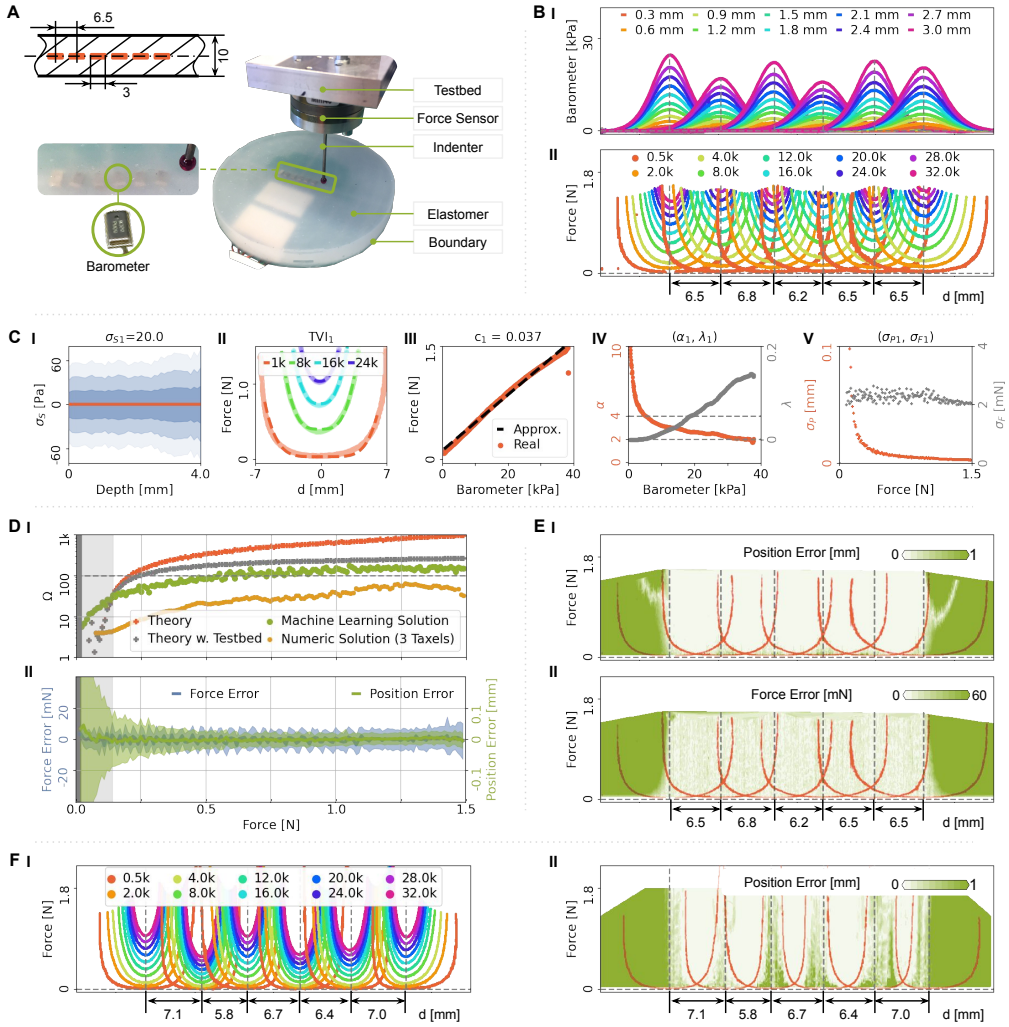

Figure 5.9: Super-resolution in 1D. Real line-sensor device with six barometer taxels at a distance of about 6.5 mm. **A**: Sensor device with stimulation testbed. **B**: Taxel response for different indentation depths (**I**) and resulting TVIs for different sensor values (**II**). **C**: Analytical calculation based on the theory performed for one of the taxels. In **I** we assure that the sensor measurement noise is constant with $\sigma_s = 20.0\,\text{Pa}$. The position and force std. dev (**V**) are analytically derived from the approximated TVIs (**II**) with appropriate parameters c, α, λ (**III, IV**). **D, E**: Quantitative evaluation of the sensor device using machine learning models to infer the position and force magnitude. **D-I**: Super-resolution factor depending on the contact force magnitude as predicted by the theory w/o and with testbed precision, and as achieved by the machine-learning solution and the numeric solution. **D-II**: Position and force error of the machine-learning solution depending on stimulation force: mean and std. dev over the 32.5 mm sensing surface between taxel 1 and 6. **E**: Spatially resolved position error (**I**) and force error (**II**). The orange lines are the TVIs for the smallest sensor value (500 Pa in **B-II**). **F**: TVIs (**I**) and spatially resolved position error (**II**) for another 1D sensor layout with varied adjunct sensor distances.

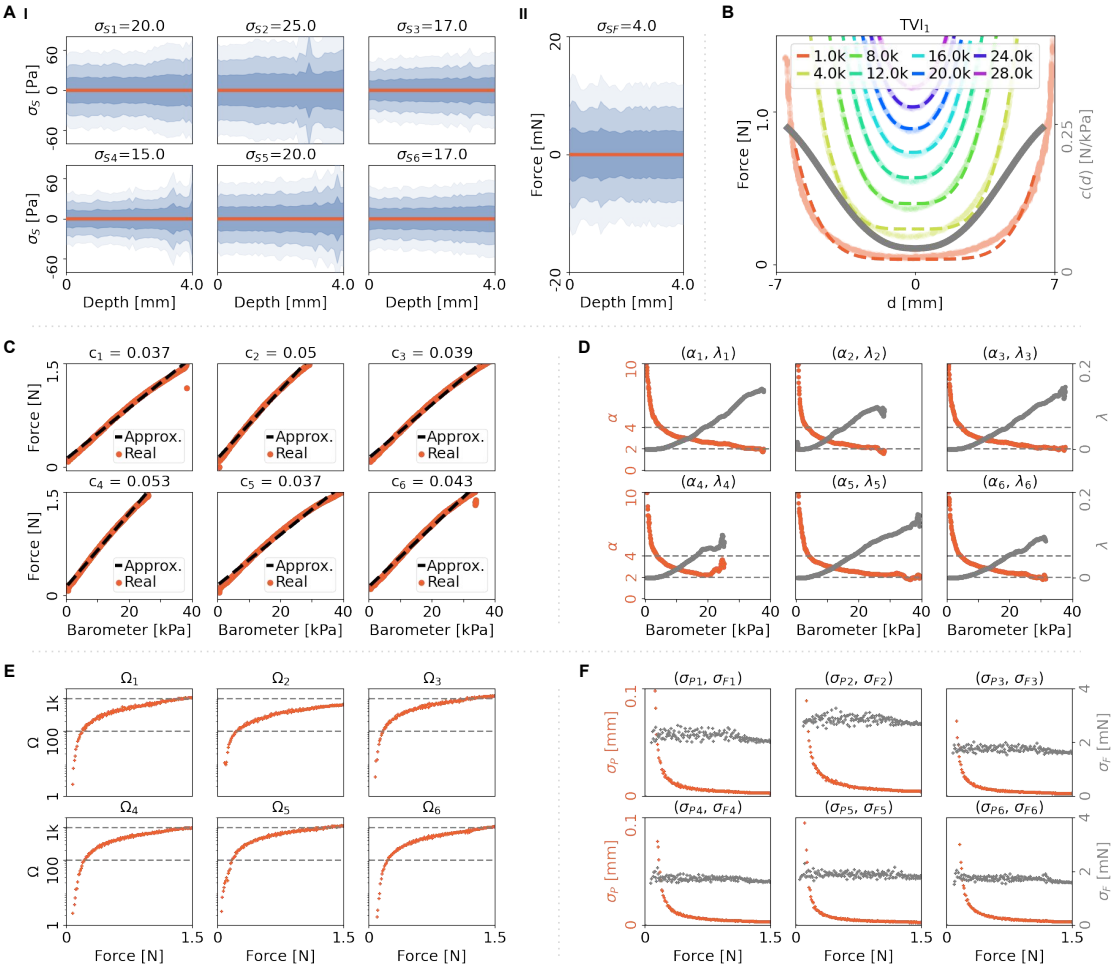

Figure 5.10: Quantitative parameters of the 1D sensor. **A** shows the measurement noises of six barometers and the force-torque sensor. **B** shows individual isolines with curved-fitted functions. $c(d)$ is a position-dependent variant that denotes the non-linear relationship between sensor values and the force values. **C**: Constant c indicates the relationship between the force values and barometer values at $d = 0$. Red is the real relationship, dashed blue is the linearly approximated relationship. **D**: Attenuation power and coefficient (α, λ) of the TVI's shape. **E**: Analytically derived super-resolution factor w.r.t. applied force strength. The dashed gray lines are with a super-resolution factor of 100 and 1 000. **F** shows the computed theoretical σ_P and σ_F w.r.t. the applied force magnitude.

Theoretical prediction

Measurement noise As mentioned above, the measurement noise (σ_S) in the sensor elements introduces uncertainties in the position and force strength inference (σ_P and σ_F). We collect 1 000 samples from the barometer and the force-torque sensor under different static loads to evaluate the measurement noise levels, as shown in Fig. 5.9 C-I and Fig. 5.10 A-I. The noise levels are almost invariant to the indentation depth (the applied force), such that we treat it as a constant.

TVI fitting and sensor-force relationship To support the assumption for our theory introduced in Section "The Model", we herein introduce the analytical steps for approximating taxel-value-isolines (TVIs) from the sensor and force measurements. We first empirically acquire the sensor isolines, and then fit a function

$$I^S(d) = g(S) + \lambda |d|^\alpha \tag{5.16}$$

to each individual isoline, as shown in Fig. 5.9 C-II and Fig. 5.10 B. Taking the $g(S)$ of each barometric values at $d = 0$ as the force strength, we can acquire the relationship between sensor value and force strength captured by the constant c (see text below Eq. 5.2). The value of c is simply the ratio between the applied force $g(S)$ and the barometer value. As shown in Fig. 5.9 C-III and Fig. 5.10 C, the linearity assumption holds for all six barometers. Here we make a surprising discovery: the attenuation coefficient λ and power α depend strongly on the barometer value (Fig. 5.9 C-IV and Fig. 5.10 D). For small values (small forces) $\alpha \approx 10$ whereas for larger values $\alpha \approx 2$. Another complication that arises from the non-linearity is that the relationship between sensor values (in kPa) and the forces values (in N) also becomes position dependent. To convert the measured σ_S into forces, we introduce

$$c(d) = \frac{\mathrm{d}F(d)}{\mathrm{d}s(d)}, \tag{5.17}$$

which denotes the derivative of the force-to-sensor-value relationship in units $\frac{N}{kPa}$. Note that $c = c(0)$.

Computation of σ_P and σ_F The impact of the attenuation power nonlinearity on the localization accuracy (σ_P) is visible in Fig. 5.3. The TVIs parameters are specific for each barometer values which reflects the non-linearity in the system. To compute the correct σ_P and σ_F we need to consider both the position d and also the contact force F, which can be understood from Fig. 5.10 B. Let i indicate which barometer's TVIs we are using,

d denote the contact position between two barometers (0–D), and F is the contact force:

$$\sigma_{Pi}(d,F) = \frac{2c_i(d)\sigma_{Si}}{\lambda_i(d,F)\alpha_i(d,F)|d|^{\alpha_i(d,F)-1} + \lambda_i(D-d,F)\alpha_i(D-d,F)|D-d|^{\alpha_i(D-d,F)-1}}, \tag{5.18}$$

$$\sigma_{Fi}(d) = 2c_i(d)\sigma_{Si}. \tag{5.19}$$

The values of $c(d)$, $\alpha(d,F)$, and $\lambda(D-d,F)$ are obtained in a numeric fashion by interpolating between the estimated TVIs displayed in Fig. 5.10 B.

As shown in Fig. 5.9 C-V and Fig. 5.10 F, placing two identical taxels at a distance of 6.5 mm and using Eq. 5.14 results in force-dependent uncertainties. The uncertainty about the force magnitude σ_F is relatively constant, whereas the localization uncertainty σ_P decreases drastically with increasing contact force. In addition to the excitation of more taxels this is a further reason for low prediction errors at higher contact forces.

Computation of super-resolution factor Ω We assume the 1D line sensor (with size 32.5 mm) has the same TVIs for all six barometers and use

$$\Omega_i(F) = \frac{32.5}{6 \cdot 2\langle \sigma_{Pi}(d,F)\rangle_d}, \tag{5.20}$$

where $\langle\cdot\rangle_d$ denotes the average w.r.t. to all postions $d \in [0,D]$, to calculate the super-resolution factor. Figure. 5.10 E shows the computed theoretical Ω w.r.t. the applied force magnitude. The overall factor Ω for the 1D case is 495 which is the average over values for each reference sensor i and over the force range (0.02–1.5 N) in discrete steps of 0.01 N: $\Omega = \langle \Omega_i(F)\rangle_{F,i}$. The red dots in Fig. 5.9 D-I show the averaged Ω over six sensors w.r.t. the contact force strength.

Data collection influence on σ_P and σ_F Our testbed has a position resolution of 0.0075 mm and the force-torque sensor has a measurement noise level (σ_{SF}) of 4.0 mN, see Fig. 5.10 A-II. The data collected by this testbed system affects the evaluation of the sensor performance. Taking the testbed into account, we denote the uncertainties as $\tilde{\sigma}_P$ and $\tilde{\sigma}_F$ and compute them as:

$$\tilde{\sigma}_{Pi}(d,F) = \sigma_{Pi}(d,F) + 0.0075, \tag{5.21}$$

$$\tilde{\sigma}_{Fi}(d) = \sigma_{Fi}(d) + \sigma_{SF}. \tag{5.22}$$

Considering the testbed effect, the overall $\tilde{\Omega}$ is 187 on average. The difference between Ω and $\tilde{\Omega}$ is visualized in Fig. 5.9 D-I and Fig. 5.11 A.

These two super-resolution factors can be taken as upper baselines to evaluate the signal processing methods.

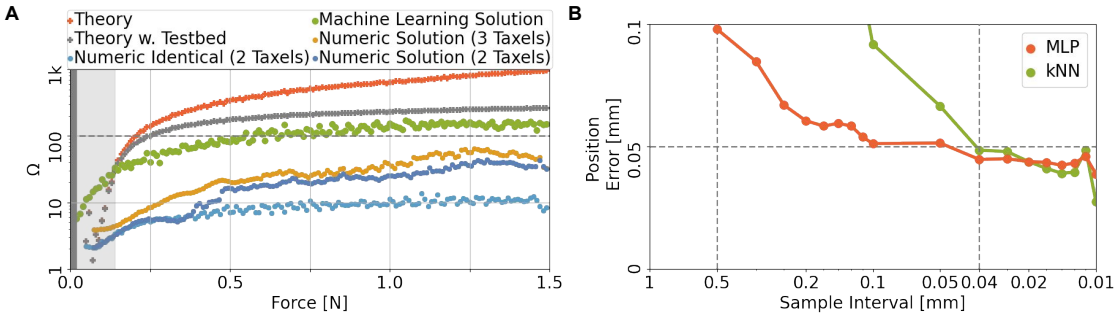

Figure 5.11: Quantitative evaluation of the 1D sensor. **A** The super-resolution factor (Ω) is acquired using different criterion: red theory baseline, gray theory baseline considering testbed effect ($\tilde{\Omega}$), green machine-learning solution, yellow numeric solution using three taxels, and blue numeric solution using two taxels. The light blue dot line offers a lower bound for the numeric solution by taking all the barometers identical and repeating the numeric solution with two taxels. **B** shows the influence of the data collection density on the position accuracy achieved by two machine learning methods: MLP and kNN. The MLP can localize contact with error lower than 0.1 mm and 0.05 mm being trained on data with 0.5 mm and 0.04 mm contact spacing, respectively.

Practical force inference

To solve the inverse problem of predicting the indentation position from the sensor measurements, we compare two methods. We first consider a numerical solution by taking three sensors with the highest sensor values among six, looking for the intersection points ($P_{1,2}$, $P_{2,3}$, $P_{1,3}$) of the TVIs, and averaging them to get the contact location. The second and preferred method for data processing is instead a neural network regression model (MLP structure) using squared error loss, which yields a prediction with minimal variance. In this way, we circumvent a manual computation of intersection locations, which suffers from inaccuracies due to real-world deviations from the idealized TVIs.

Numeric Calculation We take two barometers (i_1 and i_2) with the highest sensor values among six, compute the intersection point (P_{i_1,i_2}) of the TVIs, and take it as the contact location. Similarly, we can take three sensors (i_1, i_2, and i_3) with the highest sensor values among six, compute the intersection points (P_{i_1,i_2}, P_{i_2,i_3}, P_{i_1,i_3}) of their TVIs, and average them to get the contact location. Naturally, if only two sensors are active we fall back to the solution for two sensors. We compare the numerical solution with the ground truth contact location and compute the super-resolution factor Ω using

$$\Omega^n(F) = \frac{32.5}{6 \cdot 2 \cdot \mathrm{RMSE}^n(F)} \tag{5.23}$$

where $\mathrm{RMSE}^n(F)$ stands for error for data points with force F in bins of 0.01 N in the range of (0.02–1.5 N) for the method with $n = 2$ or $n = 3$ taxels. The resulting numbers are compared to the other methods in Figure. 5.11 A. As expected, using the TVIs from one of the sensors is worse than using all 6 and using three taxels is better than two. Nevertheless, the numeric methods fall behind the machine learning method.

Machine Learning We use a standard MLP (multi-layer-perceptron) with six fully connected hidden layers with 100 neurons each. The data consist of 132 k samples that are split into datasets of training, validation, and test with a ratio of 3:1:1 according to locations. All locations are in the 32.5 mm range between the two outermost taxels. This results in an average spacing between locations of 0.017 mm in the training data set. We compare the machine-learning solution with the ground truth contact location and compute the super-resolution factor Ω using Eq. 5.23. The result is displayed in Fig. 5.11 A.

As summarized in Fig. 5.9 D & E. We take the root-mean-square error (RMSE) as σ_P to calculate the super-resolution factor (Eq. 5.4), shown as yellow dots for the numerical solution and green dots for the machine-learning solution in Fig. 5.9D-I. The inference accuracy of the position is higher with stronger indentation forces. The machine-learning solution approaches the theoretical prediction (including the testbed errors) and outperforms the numerical solution (3 taxels). As expected, the theoretical predictions are higher, as they represent the best achievable results. Noteworthy, the machine-learning solution nevertheless excels the theory in the low force operation range (0.02–0.14 N). An explanation is that in this region the force is below the minimal force (sensitivity) for many contact locations (1 kPa shown in Fig. 5.2 C-II and Fig. 5.9 C-II) where our theory does not make predictions as no TVI intersections are available. The machine-learning solution can still predict the taxel center in these cases. Numeric values for localization errors of all methods are provided in Table 5.2.

The inference accuracy of the position using MLP is higher with stronger indentation force and slightly lower for force magnitude, but generally very accurate (evaluated at locations that were not included during training), as shown in Fig. 5.9 D-II. Averaging it over the applied force range (from 0.02 N to 1.5 N), we obtain an average super-resolution factor of 106 with the ML method, compared to the prediction of 187 from the theory with testbed consideration.

The spatial distribution of position error and force magnitude error of the prediction models are shown in Fig. 5.9E. To show the impact of different overlappings of the TVIs, we present in Fig. 5.9 F-I a sensor with varying distances between the sensor elements. The position errors (Figure. 5.9 F-II) increase between taxels with larger distances. In both cases, the overall shape resembles our theoretical prediction, cf., Fig. 5.2: higher errors occur in locations where fewer TVIs overlap. Note that this is not a problem of lacking data, as we record plenty of data also in these regions.

Our analysis uses data with a very fine spacing of contact locations (every 0.01 mm).

However, how much data is really required? We investigate the performance of the machine learning inference method depending on the spacing of training contact points and present the result in Fig. 5.11 B. Already with a contact interval of 0.5 mm a localization accuracy below 0.1 mm is achieved (evaluated on unseen contact locations at 0.01 mm resolution in a force range of 0.02–1.5 N. For comparison, we also consider a simple k-Nearest-Neighbors (kNN) method, mimicking a lookup table. For coarse data the kNN falls short, but for data finer than 0.04 mm resolution both methods perform similarly.

5.7.2 Quantitative Analysis in 2D

Table 5.2: In 1D we compare the theoretical predictions with the machine learning (ML) approach and our numerical computation. For the 2D case we report, in addition to the used MLP method, a version with separate prediction for x and y component with increased performance.

Force Interval [N]	0.02–0.14	0.14–0.2	0.2–0.4	0.4–0.6	0.6–0.9	0.9–1.2	1.2–1.5	overall
1D Theory	0.714	0.052	0.017	0.008	0.005	0.004	0.003	0.053
1D Theory + Testbed	0.721	0.060	0.024	0.016	0.013	0.012	0.011	0.060
1D ML	0.196	0.079	0.051	0.032	0.024	0.019	0.018	0.042
1D Numeric (3 Taxels)	0.668	0.517	0.264	0.134	0.098	0.068	0.049	0.158
1D Numeric (2 Taxels)	1.074	0.678	0.476	0.203	0.129	0.103	0.071	0.244
2D ML	0.567	0.230	0.158	0.120	0.107	0.107	0.120	0.161
2D ML (separate x,y)	0.485	0.219	0.139	0.101	0.086	0.082	0.082	0.132

The above analysis for the 1D case helps us to investigate a sensor with a flat or curved 2D sensing surface. We mold 25 barometers in the elastomer (Dimension: 34 mm $\times$ 34 mm, Thickness: 10 mm) with a 5 $\times$ 5 grid layout with approximately 6.5 mm distance to each other as shown in Fig. 5.12 B-I. A testbed carries a 4 mm spherical indenter and contacts the surface at given locations, similar to Fig. 5.9 A. Sensor values, force values, and indentation positions and depths are recorded at 69 $\times$ 69 positions (a grid with 0.5 mm apart from each other) evenly distributed on the sensing surface (34 mm $\times$ 34 mm) with 20 incremental indentation depths (0.2 mm each) at each position. The resulting taxel responses are presented in Fig. 5.13. We employ two machine learning models to predict the indentation position and force magnitude, respectively. We use an MLP with ten fully connected hidden layers with 100 neurons each. The data consists of 95 k samples that are split into datasets of training, validation, and test with a ratio of 3:1:1. The models are trained with the *L2* loss, the Adam optimizer (learning rate: $2 \cdot 10^{-4}$, epsilon: 10^{-5}), and a batch size of 100 samples in 1 million iterations. The models for position and force inferences are separately trained using the same architecture and training settings.

Using the trained machine-learning models, we evaluate the super-resolution factor of the 2D sensor as the fraction between the number of virtual taxels n_v and the real number

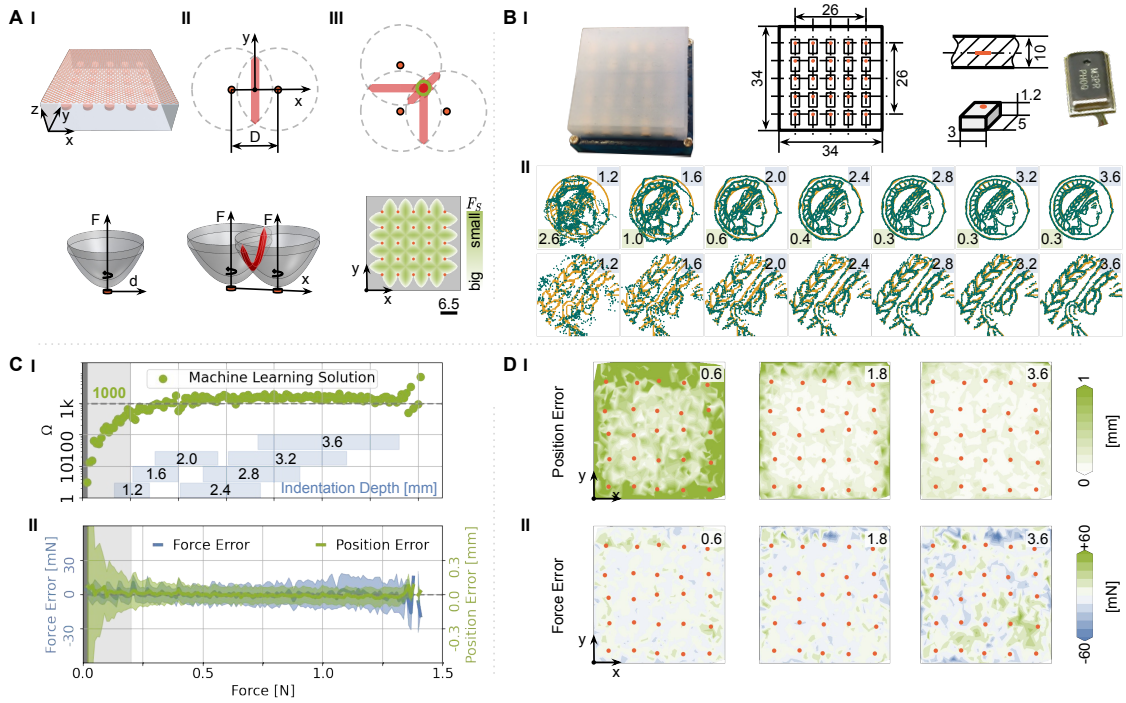

Figure 5.12: Super-resolution in 2D. **A**: 2D sensor arrangement and taxel values isolines for a 2D sensing surface. **A-I bottom**: Taxel value isoline (iso-surface) for a single taxel. **A-II**: Intersection volume (due to measurement uncertainty) for two taxels at a distance D along the x axis, see top view. The localization would be very uncertain along the y-direction. **A-III**: Proper localization requires at least three taxels for one contact point. The lower part shows a grid sensor arrangement with resulting sensitivity over the surface. Locations between taxels are more sensitive, requiring a smaller force (F_S) to activate them. **B** presents our custom 2D sensor designed according to our theory guidelines. **B-I** is a picture of the sensor together with its geometric properties in a millimeter (mm) scale. **B-II** shows a performance illustration for the super-resolution functionality. A 4 mm spherical indenter contacts the surface following a circular "Minerva" pattern (yellow) with different indentation depths (stated in the upper right corner), and a trained machine learning model infers contact locations (green) with RMSE [mm] indicated in the bottom left corner. The lower row shows a detailed view of the central region. **C, D**: Quantitative evaluation of the sensor device using machine learning models to infer the position and force magnitude. **C-I**: Super-resolution factor depending on contact force magnitudes over the sensor's center region (26 mm $\times$ 26 mm) **C-II**: Position and force errors depending on stimulation forces: mean and std. dev. **D**: Spatially resolved position error (**I**) and force error (**II**) w.r.t. indentation depths.

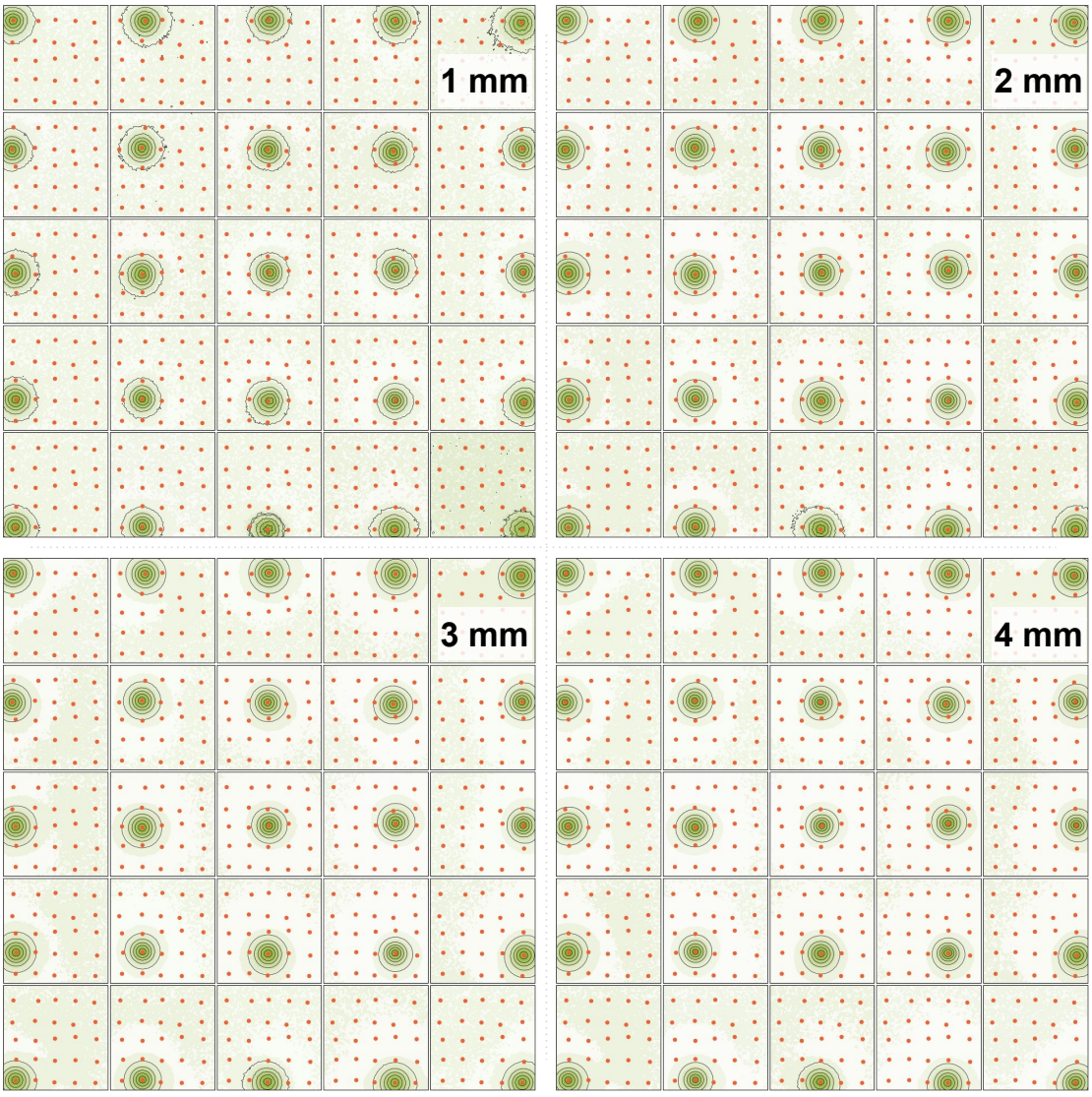

Figure 5.13: Response and TVIs for real sensor elements on a 2D sensing surface. 25 barometer values for 2D sensor with different indentation depths (1 mm, 2 mm, 3 mm, 4 mm). Each sub-figure with the specific indention depth shows the barometer value contour maps for each of the 25 sensors marked by a red dots (center position).

of sensors (taxels) $n_r = 25$ with the following equation:

$$\Omega(F) = \frac{n_v(F)}{n_r} = \frac{A/A_v(F)}{n_r} = \frac{A}{n_r \cdot \pi \cdot \mathrm{RMSE}_x(F) \cdot \mathrm{RMSE}_y(F)}, \tag{5.24}$$

where $A = 26\,\mathrm{mm} \cdot 26\,\mathrm{mm}$ is the center sensing area of interest, and $A_v(F)$ is the area of a virtual taxel calculated as an ellipse with radii of the standard deviations (RMSE)s in x and y directions.

The results are summarized in Fig. 5.12 C & D. Figure. 5.12 C-I shows the averaged super-resolution factor of 1260 using

$$\Omega_k = \frac{A}{n \cdot \pi \cdot \mathrm{RMSE}_{x,k} \cdot \mathrm{RMSE}_{y,k}} = \frac{26 \cdot 26}{25 \cdot \pi \cdot \mathrm{RMSE}_{x,k} \cdot \mathrm{RMSE}_{y,k}}, \tag{5.25}$$

where A is the sensing area and k is the k^{th} force interval (0.01 N) over a force range of 0.02 to 1.5 N. The inference accuracy of the position is higher with stronger indentation force, as shown in Fig. 5.12 C-II. The force accuracy is relatively constant, with 0.01 N (RMSE) on average. The averaged error of inferred locations is at 0.161 mm (RMSE) and for forces larger than 0.2 N an error 0.120 mm (RMSE) is achieved. The spatial distribution of position and force magnitude error for different indentation depths are shown in Fig. 5.12 D. With increased indentation depth, the position accuracy improves. In coherence with our theory, at the boundaries this improvement is smaller because less TVIs overlap. In comparison, the spatial distribution of the force accuracy is relatively homogeneous. The error increases slightly with increased indentation depth, and forces at the boundaries tend to be slightly underestimated. Interestingly, the performance can still be improved by training MLP models for x and y separately, see Table 5.2).

5.7.3 Quantitative Analysis in 3D

We exploit the above analysis to design a 3D, dome-shaped to facilitate real applications. We first validate the influence of the sensor placement and the thickness of the elastomer on the TVIs. As shown in Fig. 5.14 A, the barometer directly fixed to the boundary has a narrower perception field and saturates more easily. The perception field doesn't change much with reduced thickness, but the force needed to activate the same sensor value is smaller than Fig. 5.8 B-III.

We mold sixteen wired barometers floating inside an elastomer with a thickness of 7.2 mm. The sensor named **BaroDome** has a dome-shaped geometry comprising a hemisphere (radius: 10 mm) and a cylinder (radius: 10 mm, height: 16 mm), as shown in Fig. 5.14B & C. We build a 5 DoF testbed to collect normal/shear quasi-static contact data over the sensing surface, as shown in Fig. 5.14B & C. Measurements are collected using our automated testbed to probe **BaroDome** in different locations. To obtain a

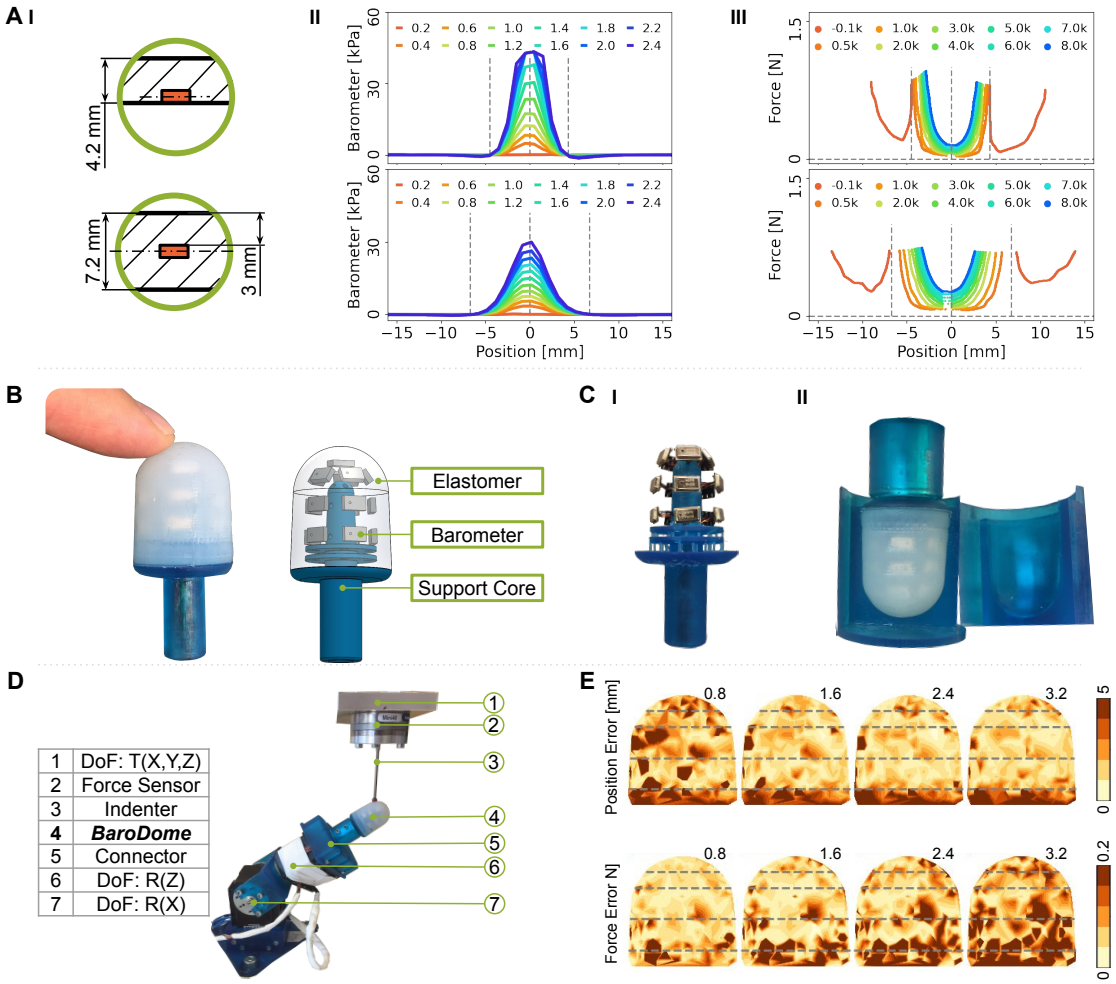

Figure 5.14: Super-resolution in 3D. **A**: Sensor placement comparison: Bonding to the boundary vs. floating in the elastomer center. **A-I** shows the geometric schematic. **A-II** and **A-III** present the spatial barometer values w.r.t. indentation depth (legend) and isolines thereof. **B** shows validation of our theory on a fingertip-sized 3D sensor. **C** presents a two-steps manufacturing procedure for the 3D sensor: wiring and molding. **D** shows a 5 DoF testbed for collecting normal and shear quasi-static contact forces over the sensing surface. **E** presents the distributions of position error and force errors w.r.t. indentation depth (indicated on each top right corner). Gray dashed lines indicate the sensor positions.

variety of normal and shear forces, the indenter is moved to a specified location, touches the outer surface, deforms it increasingly by moving normal to the surface with fixed steps of 0.2 mm (up to 3.2 mm). For each indentation level, the indenter also moves sideways to apply shear forces (normal/shear movement ratio 2:1). After a pause of two seconds to allow transients to dissipate, we simultaneously record the contact location, the indenter contact force vector from the testbed's force sensor, and the barometer values from **BaroDome**.

We employ two machine learning models to predict the indentation position and force magnitude, respectively. We use an MLP with ten fully connected hidden layers with 100 neurons each. The data consists of 110 877 samples that are split into datasets of training, validation, and test with a ratio of 3:1:1. The models are trained with the *L2* loss, the Adam optimizer (learning rate: $2 \cdot 10^{-4}$, epsilon: 10^{-5}), and a batch size of 200 samples in 1 million iterations. The models for position and force inferences are separately trained using the same architecture and training settings.

The trained machine learning models offer a median position accuracy of 1.5 mm and a median force accuracy of 0.04 N over a force range of 0.02-2.0 N. As shown in Fig. 5.14 E, localization accuracy increases when the indentation depth is more significant while the force accuracy worsens. Notice, higher force error tends to locate near the barometer measurement elements, which might be due to the sensor saturation effect.

5.8 Discussion

We present a new way to characterize, analyze, and predict force sensation at super-resolution for tactile sensors. Our theory is based on sensor isolines that allow for a direct assessment of the uniqueness of contact position reconstruction. We derive quantities such as minimum force sensitivity, localization and force sensing accuracy. These allow us to analytically compute the super-resolution factor, namely, the number of distinguishable locations between two real sensor elements and answer the initially posed questions on the obtainable super-resolution and the sensitivity.

With the help of an FEM model, we give guidelines for common design choices, such as placement of the sensor elements within the elastomer as well as material properties. We analyze three commonly used sensor types: strain gauges, accelerometers, and barometers. Both, accelerometers and barometers, show the necessary properties for super-resolution sensation within our framework. We conduct two case-studies to evaluate our theory using a line (1D) and a grid (2D) of barometer sensor elements embedded in elastomer skin. Following our theory-informed design choices, the 1D-sensor has a theoretically predicted super-resolution factor of 187. Using machine learning as a practical implementation achieves a remarkable performance of 106-fold super-resolution. Our 2D sensor with 26-26 mm sensing surface reaches a 1 260-fold super-resolution and can localize contacts up to 0.161 mm on average.

We hope that our approach can help the design of new and capable tactile sensors.

The major insights from our study regarding the question of soft material properties and sensing method can be summarized as follows. The sensor elements in the transmission medium (elastomer) should have convex isolines. It is beneficial to have the sensor elements "float" in the center of the elastomer or closer to the sensing surface. Flexible wiring helps to have the best sensitivity. A thicker elastomer layer seems beneficial, and materials with small Young's modulus, high Poisson's ratio, and big yield strength are recommended. The distance between sensor elements should be such that for a majority of forces, the isolines of neighboring taxels intersect. For inferring simultaneous contacts, a single sensor element should be only activated by a single contact, and the distinguishable distance increases when the force magnitudes of multiple contacts are higher.

The decisive quantity in our theory for super-resolution perception is spatial distribution of the uncertainty of contact location and force magnitude inference. This uncertainty can be decreased by averaging subsequent samples from the sensors (thus increasing the response time) and by moving the taxels closer together to excite more than two taxels per contact. In a practical implementation, we need a mechanism to perform the actual tactile information inference. This can be implemented numerically and using machine learning, where the latter shows a 3 times better performance. How is the connection between the ML model and the theory? Exploiting the equivalence of minimizing model uncertainty by maximum likelihood estimation and least-squares error minimization for Gaussian residuals, the machine learning model directly optimizes for the quantities of interest: the prediction uncertainties. Whereas the theory describes an upper limit for the super-resolution capabilities under the model-assumptions, we find that the ML methods approach the predicted accuracy in the considered real-world sensors. In this way, our theory offers guidelines towards the system-level design of machine-learning-driven tactile sensors.

Besides the above-mentioned mechano-electrical properties, the predictions by the theory can be used to validate the suitability of the employed ML model, the data collection, and the training procedure, as we can expect performances to come close to the predicted ones. We show how the inaccuracies in the testbed need to be taken into account and how much data is effectively required for good performance.

Our work enables engineers to make more informed decisions when aiming for high-resolution tactile sensing. New designs can be invented that create more robust and cheaper sensors without sacrificing the required precision. The principles employed in this paper are applicable to a larger range of sensing mechanisms. Optimizing a design for accelerometers or hall-effect sensors might be a promising next step. We hope our work delivers an important stepping stone to ubiquitous tactile sensing in robotics. We provide further evidence that machine learning methods are a flexible way of performing data processing in tactile sensing. They are capable of approaching the theoretical limit in terms of super-resolution sensing, provided a sufficient amount of high-quality data is available. Promising directions for future research include the extending our work to shear forces and the investigation of structured transmission media e.g., with ridges [170] or multiple layers [10].

Chapter 6

Insight: a Haptic Sensor Powered by Vision and Machine Learning

In this chapter, we present a robust, soft, low-cost, vision-based, thumb-sized 3D haptic sensor named *Insight*: it continually supplies the host robot with a directional force-distribution map over its entire conical sensing surface. *Insight* uses an internal monocular camera, photometric stereo, and structured light to detect the 3D deformation of the easily replaceable flexible outer shell, which is molded in a single layer over a stiff frame to guarantee sensitivity, robustness, and a soft contact surface. The force information is computed by a deep-neural-network-based machine-learning method that maps images to the spatial distribution of 3D contact force (normal and shear), including numerous distinct contacts with widely varying contact areas. Extensive experiments show that *Insight* has an overall spatial resolution of 0.4 mm, force magnitude accuracy around 0.03 N, and force direction accuracy around 5 degrees over a range of 0.03–2 N. It is sensitive enough to feel its own orientation relative to gravity, and its tactile fovea can be used to sense object shapes. The presented hardware and software design concepts can be extended to achieve robust and usable tactile sensing on a wide variety of robot parts with different shapes and sensing requirements.

6.1 Introduction

Many efforts have been made to create haptic sensors that can quantify contacts across the surface of a robot [5]. Previous successful designs produced measurements using resistive [6, 7, 8, 9], capacitive [10, 11, 12], ferroelectric [13], triboelectric [14], Hall-effect-based [15] and optoresistive [16, 17] transduction approaches. Vision-based haptic sensors [18, 19, 20, 21, 22, 23] have demonstrated a new family of solutions, typically using an internal camera that views the soft contact surface from within.

A detailed comparison of representative state-of-the-art sensors is shown in Table 6.1. The mechanical designs of all previously developed sensors employ multiple functional layers, which are complex to fabricate and can be delicate. Many tasks benefit from a large 3D sensing surface rather than small 2D sensing patches; however, only a few sensors possess a 3D surface [22, 108, 107]. Their design is often technically complex,

Table 6.1: A comparison between state-of-the-art haptic sensors and our design. An upward arrow (↑) indicates that higher values are better, while a downward arrow (↓) means lower is better.

Sensor Name	Transduction Method	# of Layers ↓	Surface Shape ↑	Area ↑ [mm^2]	A/V ↑ [mm^{-1}]	Data Processing	Output Format	Sensing Error ↓ P [mm]	F_n [N]	F_s [N]	α [°]	Notes
BioTac	Resistive	2	Half 3D	484	0.060	FCN	Location+Force	1.4 [50]	0.85	∼0.48 [6]	10 [51]	costly and delicate
Lee et al. [173]	Capacitive	5	2D	484	1.000	—	Location	2	—	—	—	10 mN full-scale range
Yan et al. [15]	Hall Effect	2	2D	324	0.179	FCN	Location+Force	0.1	0.15	—	—	blind to shear force
SLIMS [17]	Optical	3	1D	350	0.050	—	Location	10	—	—	—	sensitive to ambient light
GelSight [18]	Cam+PS+Markers	5	2D	250	0.003	CNN	Shape+Force	—	0.67	∼0.17	—	complex to manufacture
GelSlim [109]	Cam+PS+Markers	6	2D	1200	0.006	inv.FEM	Force Map	—	0.32	∼0.22	—	tears after 1500 contacts
OmniTact [108]	5×Cam+PS	2	3D	3110	0.083	ResNet	Location	0.4	—	—	—	5 cameras, ambient light
GelTip [107]	Cam+PS	3	3D	2513	0.052	Numeric	Location	5	—	—	—	imaging artifacts on tip
TacTip [22]	Cam+Markers	3	3D	2500	0.025	Numeric	Location	0.2	—	—	—	marker density limits res.
Sf. & D'A. [20]	Cam+Beads	3	2D	900	0.008	FCN	Location+Force	0.2	0.05	—	—	heavy, blind to shear
Sf. & D'A. [175]	Cam+Beads	3	2D	900	0.008	CNN	Force Map	—	0.13	0.05	—	no real shear experiment
DIGIT [174]	Cam+PS+Markers	3	2D	304	0.031	ResNet	Shape	—	—	—	—	difficult to extend to 3D
Insight	**Cam+PS+SL**	**1**	**3D**	**4800**	**0.088**	**ResNet**	**Force Map**	**0.4**	**0.03**	**0.03**	**5**	**this paper**

for instance, with multiple cameras [108] or special lenses [22]. Depending on their mechanical design, sensors also have widely varying sensing surface area and sensor volume. We provide *area per volume* (A/V) in Table 6.1 as a measure of compactness. Existing sensors also differ in the type of information they provide. Most sensors provide only localization of a single contact [173, 17, 108, 107, 22], and some sensors additionally provide a force magnitude [6, 15, 20] without force direction. Others are specialized for measuring contact area shape [18, 174]. Although in reality contacts are often multiple and complex, a spatially extended map of 3D contact forces over the surface, which we call a *force map*, is only rarely provided [109]. Many sensors rely on analytical data processing [107, 22, 49, 19], which requires careful calibration. It is difficult to obtain correct force amplitudes with such an approach because materials are often inhomogeneous and the assumption of linearity between deformation and force is often violated. Data-driven approaches, as used in [6, 18, 108], can deal with these problems but naturally require copious good data.

In this chapter, we present a new soft thumb-sized sensor with all-around force-sensing capabilities enabled by vision and machine learning; it is durable, compact, sensitive, accurate, and affordable (less than 100 USD). We name our sensor *Insight* as it consists of a flexible shell around a vision sensor. We initially designed the sensor for dexterous manipulation devices with behavioral learning scenarios in mind. However, our sensor is adequate for many other applications, and our technology can be adapted to create a variety of differently shaped 3D haptic sensing systems.

Figure 6.1 shows the principles behind the design of *Insight*. The skin is made of a **soft elastomer** over-molded [133] on a hollow **stiff skeleton** to maintain the sensor's shape and allow for high interaction forces without damage (Fig. 6.1B). It utilizes **shading** effects [176] and **structured light** [177] to monitor the 3D deformation of the sensing surface with a single **camera** from the inside (Fig. 6.1C). The sensor's output is computed by a data-driven **machine-learning** approach [129, 9], which directly infers distributed contact-force information from raw camera readings, avoiding complicated calibration or

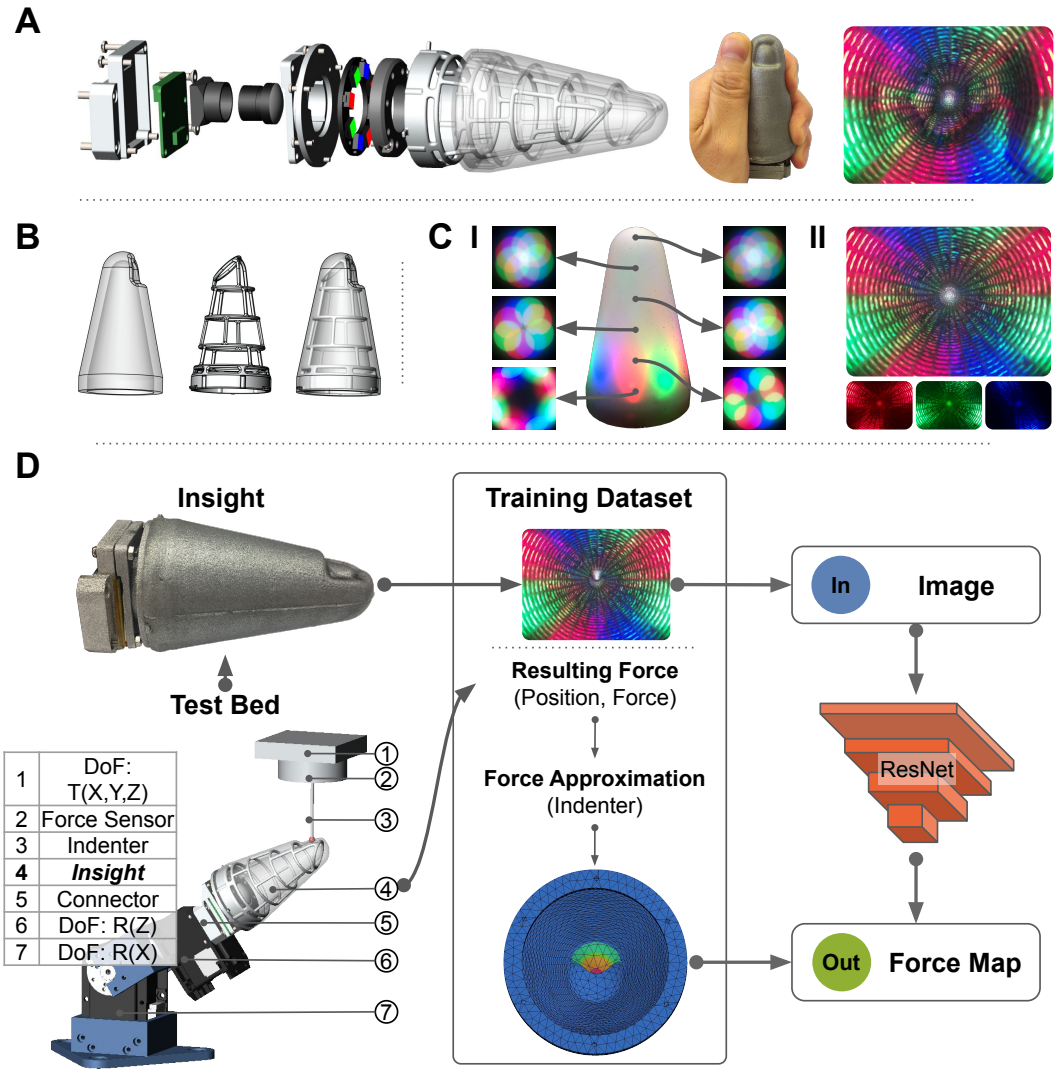

Figure 6.1: **The design principles of *Insight*.** **A** depicts the overall structure of the sensor with its hybrid mechanical construction and internal imaging system. For comparison, the sensor is shown in a human hand, next to the corresponding camera view. **B** shows the pure elastomer (left), the stiff hollow skeleton (middle), and both over-molded together (right). **C-I** illustrates the internal lighting using a translucent shell: the LED ring with apertures creates light cones, visualized by their projections on flat horizontal planes. **C-II** depicts light projection patterns within as seen by the camera in the undeformed opaque shell. **D** presents the data processing pipeline. The machine-learning model is trained on data collected by an automatic test bed. Each data point combines one image from the camera with the indenter's contact position and orientation, contact force vector, and diameter, which are used to calculate a ground-truth force distribution map from an approximate model under consistent contact forces.

any hand-crafted post-processing steps (Fig. 6.1D).

Insight is evaluated against several rigorous performance criteria within this chapter. When indented by a hemispherical tip with a diameter of 4 mm at a force amplitude up to 2.0 N, the sensor can achieve an average localization accuracy around 0.4 mm and a force accuracy around 0.03 N. By directly estimating both normal and shear components of each applied force vector, the sensor reaches an average directional estimation error around 5°. Moreover, in the absence of contact, *Insight* is sensitive enough to recognize its posture relative to gravity based only on the deformations caused by its own self-weight, which are not detectable by eye.

This chapter is structured as follows: We start with the description of the sensor operation and design principles in Section 6.2, and introduce the fabrication and characterization of the sensor in Section 6.3. In Section 6.4, we evaluate the sensor performance from different perspectives. We furthermore compare our sensor design with other state-of-the-art sensors in Section 6.5.

6.2 Principles of Operation and Design

At the core of our design is a single camera that observes the sensor's over-molded elastic shell from the inside (Fig. 6.1A). Photometric effects and structured lighting enable it to detect the tiny deformations of the sensor surface that are caused by physical contact. In principle, the contact force vectors could be numerically computed from the observed deformations according to elastic theory, but the material properties are not uniform, and the necessary assumption of a linear relationship between deformation and force [49, 178] is often violated. In our approach, machine learning greatly simplifies this process by mapping images directly to force distribution maps. The details are shown in Fig. 6.1 and explained in the following.

6.2.1 Mechanics

We aim at a compliant and sensitive sensing surface because of the favorable properties of soft materials for manipulating objects [35], for safer interactions in human environments [36], and to limit the instantaneous impact forces caused by unforeseen collisions in robotic systems [37]. Nevertheless, the direct application of soft materials in sensor design is nontrivial because they cannot withstand larger interaction forces. If thin structures are formed from a material with low Young's modulus, even gravity and inertial effects change their shape considerably [179].

To ensure a compliant sensing surface, high contact sensitivity, and robustness against self-motion, we design a soft-stiff hybrid structure using over-molding (Fig. 6.1B) [133]. The structure is composed of two parts. One is a flexible elastomer (Young's modulus around 70 kPa, hardness around $00-30$ in Shore-00) to sense the contact, and the other is a skeleton made of aluminum alloy (Young's modulus around 70 GPa) to support

the sensing surface. In this way, the sensor is not only structurally stable so it keeps its overall shape under high contact forces but also sensitive so that gentle interaction forces cause local deformations. Moreover, the sensor's shell is hollow so that the entire system is lightweight; avoiding direct contact with any optical elements also reduces the chances of image distortion and system damage. Constructing a single elastomer layer that serves all purposes is a simple, compact, robust, and wireless solution for haptic sensing. For demonstration, we include a thin, flat area of elastomer near the sensor's end for higher-resolution perception of detailed shapes (akin to a tactile fovea).

6.2.2 Imaging

Two main techniques can be used to obtain 3D information from a single camera. Photometric stereo (PS) [176] uses multiple images of the same scene with varying disparate light sources from different illumination directions to infer the 3D shape from shading information. Structured light (SL) [177] is a single-shot 3D surface-reconstruction technique that uses a unique light pattern and the fact that its appearance depends on the shape of the 3D surface on which it is projected. Generally, PS is better at capturing local details, while SL is used for coarser global reconstruction [180, 181]. PS is most effective when the illumination is nearly parallel to the surface, where the normal vectors of the deformed surface can be finely reconstructed from shading information [18]. SL allows for more perpendicular lighting of the surface and improves with larger disparity between light source and camera.

Insight is the first haptic sensor that combines PS and SL to detect the deformation of a full 3D cone-shaped surface in the single-camera single-image setting. LED light sources around the camera produce distinct light cones (eight in our prototype, as shown in Fig. 6.1C). The lighting direction is adjusted through a collimator to introduce a suitable SL pattern that favors locally parallel lighting for PS, as depicted in Fig. 6.4C. When an area of the sensor surface is contacted from the outside, the surface orientation changes, which causes a difference in color intensity through shading. The surface displacement additionally changes the distance of the surface to the camera, which can be detected with SL cones due to the color change per pixel.

6.2.3 Information

Sensors can capture many different types of haptic information, such as vibration [46], deformation [9, 47, 18], undirected pressure distribution [7], and directional force distribution [48, 49]. For robotics applications, a directional force distribution is the preferred form of contact information, as it describes the location and size of each contact region, as well as the local loading in the normal and shear directions [38]. Our proposed sensor is the first sensor designed to deliver this type of contact information, i.e., a 3D directional force distribution over a 3D conical sensing surface represented by a fine mesh of points, where each point has three force elements that are orthogonal to one another.

In a classical estimation chain, the force distribution is inferred from the surface displacement using a linear stiffness matrix based on elastic theory [178]. The displacement map can be acquired by analytically reconstructing the current normals of the sensing surface or numerically deriving the relative movement of labeled markers from the raw image captured by the camera, as done in [48, 49]. However, large deformations violate linearity between displacement and force. In addition, the over-molding in our design creates an inhomogeneous surface, where the stiffness matrix is difficult to model accurately. Shear forces are visible as small lateral deformations that highly depend on the distance to the stiff skeleton. Moreover, the reconstruction of surface normals requires evenly distributed light, without shadows or internal reflections [18]. Tracking markers [20, 175, 49] rather than a surface does not solve the fundamental problems with displacement-focused approaches.

Thus, we employ a data-driven method to estimate the force distribution directly from the raw image input using machine-learning techniques, namely an adapted ResNet [71], which is a favored deep convolutional-neural-network architecture. To collect reference data to train the neural network, we built a position-controlled 5-DoF test bed with an indenter that probes the designed sensor. A 6-DoF force-torque sensor (ATI Mini40) measures the force vector applied to the indenter so that we can simultaneously record *ground truth* forces and corresponding images from the camera inside the sensor. The target force distribution map corresponding to each contact is computed by a simple spatial approximation using the known force vector, contact location, and indenter diameter. The approximation was chosen from a set of five candidates based on the resulting contact inference performance. A subset of all data is used to train the machine-learning model. The entire process is illustrated in Fig. 6.1D and detailed in Section 6.4.6.

6.3 Fabrication and Characterization

As depicted in Fig. 6.2, the fabrication process of *Insight* includes three main aspects: the imaging system, mechanical components, and optical properties.

6.3.1 Imaging system

The sensing surface for our prototype is conical with a base diameter of 40 mm, a height of 70 mm, and a hemispherical tip. The camera's field-of-view (FoV) is adjusted to see the sensing surface with a fisheye lens: $123.8^\circ \times 91.0^\circ$ (Fig. 6.3 and Table 6.2). A ring with eight programmable tri-color LEDs is chosen for the light source. A collimator is designed and 3D printed to constrain the light-emitting path and construct the structured light pattern based on an analysis of light attenuation curves (Fig. 6.4).

Sensor geometry and FoV of the camera The sensor is designed with robotic manipulation platforms in mind, such as the TriFinger manipulator [182] illustrated in

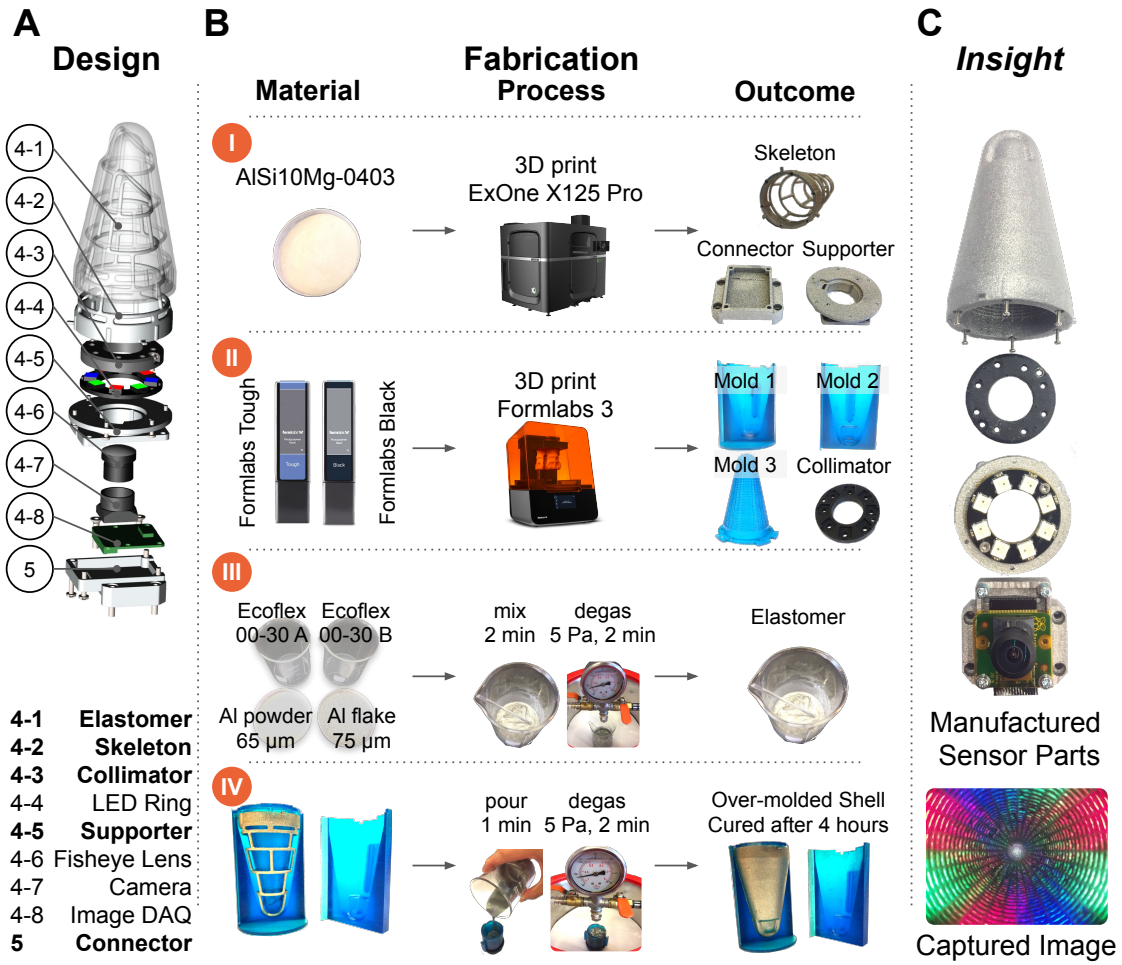

Figure 6.2: **The fabrication process of *Insight***. **A** provides an exploded view of *Insight* with all parts in the design; bold items were custom-fabricated. **B** shows the materials, processing steps, and intermediate outcomes for the elastomer, the skeleton frame, the molds for over-molding, the collimator, and the over-molded sensing surface. **C** presents the partially assembled *Insight* and an image captured under no contact.

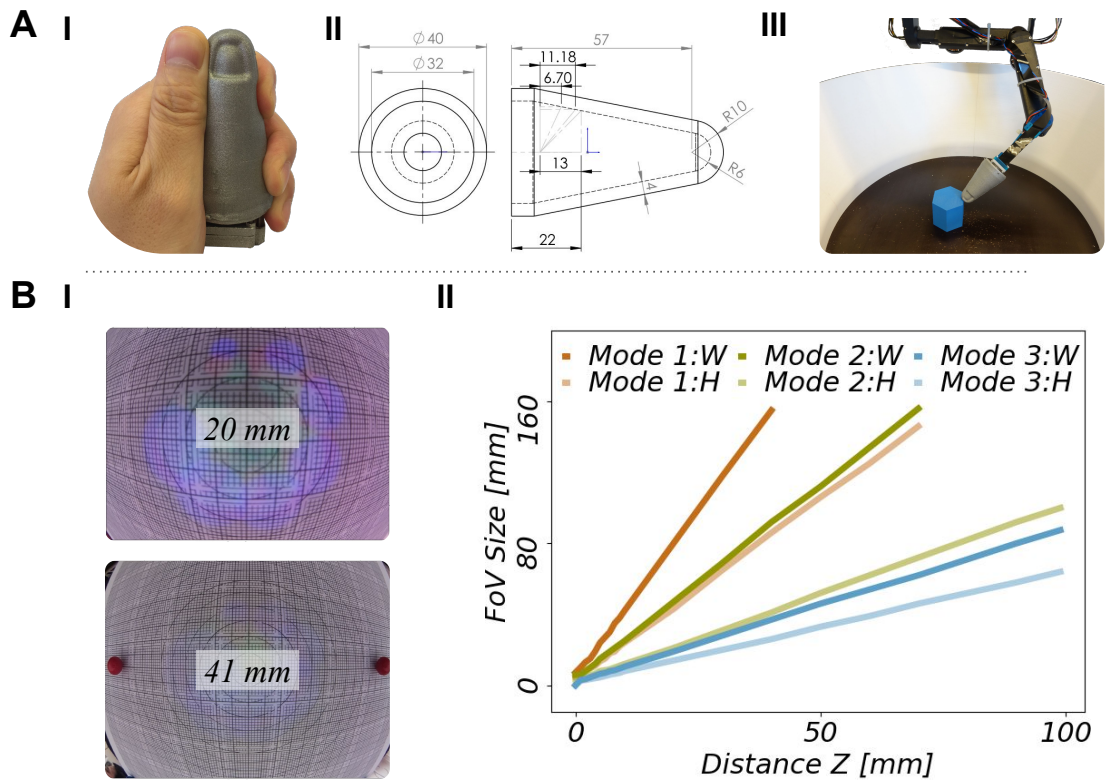

Figure 6.3: **Sensor geometry and camera FoV**. **A** shows how the sensor shape and size compare to a human thumb (**A-I**), the sensor's detailed geometry (**A-II**), and one of its intended application scenarios (**A-III**); *Insight* is designed to provide haptic sensing at the tip of the depicted robot arm to facilitate dexterous manipulation. **B** summarizes the FoV of the camera with a fisheye lens in different operating modes. **B-I** presents views of a measurement grid at two distances from the camera. **B-II** shows the width and height of the camera's FoV in the three operating modes presented in Table 6.2.

Fig. 6.3A-III. The geometry (Fig. 6.3A-II) is a cone shape with a maximal diameter of 40 mm and a height of 70 mm, similar to a human thumb (Fig. 6.3A-I). This design successfully achieves distributed 3D haptic sensing within the camera's field-of-view (FoV). The sensor can be adapted to other applications by changing its size, shape, and electronics. The following sections describe the requirements and design decisions regarding the imaging system.

To offer all-around sensing over *Insight*'s 3D curved surface, the internal camera needs to see as much of the inner surface as possible. Thus, we mount a fisheye lens with a wide-angle FoV (160°). The camera can operate in different modes with different frame rates (Table 6.2). We have chosen to operate the camera in *Mode 1*, mainly to maintain maximal FoV (Fig. 6.3B-I). However, the camera does not have equal viewing capabilities in the horizontal and vertical directions because its imaging sensor is rectangular, as

Table 6.2: FoV of the camera with a fisheye lens in different operating modes.

Mode	Image Resolution [W × H × fps]	FoV in W [°]	FoV in H [°]
1	**1640 × 1232 × 40**	**123.8**	**91.0**
2	1280 × 720 × 90	94.7	52.7
3	640 × 480 × 90	46.8	34.8

shown in Fig. 6.3B-II and Table 6.2.

Lighting system To construct the structured light pattern, we analyze the color and brightness of each light source, the behaviors of light intensity attenuation, and the parameters of the collimator as well as the camera's sensitivity to differently colored light brightness (Fig. 6.4). The light sources are generated from an LED ring with eight elements, each of which has three programmable channels to create red, green, and blue light. These elements emit light in all directions of the half 3D space and light up the near field without visible differences. We use a collimator to constrain the emitting path and construct a particular light cone for each LED, as shown in Fig. 6.4A-I and Fig. 6.4C. The collimator is optimized to create a lighting pattern where most of the surface is lit by at least two and at most four LEDs. This design enables excellent detection of deformations from the shading effects, and we avoid both over-saturated and under-lit areas.

The collimator is made from an opaque material where the holes have two key geometric parameters; see Fig. 6.4C-I. First, the diameter of the collimator D constrains the light cone size (Fig. 6.4A-II & III). Second, the tilt angle α slants the light cone radially outward, as shown in Fig. 6.4C-I & II. We measure the effect of the collimator diameter on the light cone size with a test setup. We use a test bed to move a marker board to defined distances Z from the camera and change the collimator size. The marker board has a fine grid on it, which is used to count the projected light cone diameter D_1 (Fig. 6.4A-II). The value of D_1 is computed using *full width at half maximum* (FWHM), as shown in Fig. 6.4B-IV. The projected light cone diameter scales linearly with the collimator diameter, as expected and shown in Fig. 6.4A-III. Based on these curves, we test the collimator angle effect. We tune the angle α (3°) and diameter D (2.5 mm) jointly to make a light pattern that fully covers the internal sensing surface of *Insight* as stated above (Fig. 6.4C-II).

We also conduct a detailed analysis of the light attenuation behaviors; see Fig. 6.4B-I–III. The total received light intensity attenuates approximately linearly as the brightness reduces, and approximately quadratically as the distance increases; these effects are shown in Fig. 6.4B-I and Fig. 6.4B-II. With increased distance, the light cone gets wider, but the portion of the reflected light beams seen by the camera gets smaller. The light also attenuates within one horizontal cross-section of the light cone (Fig. 6.4B-IV). We use the FWHM criterion to calculate the size of the projected light cones.

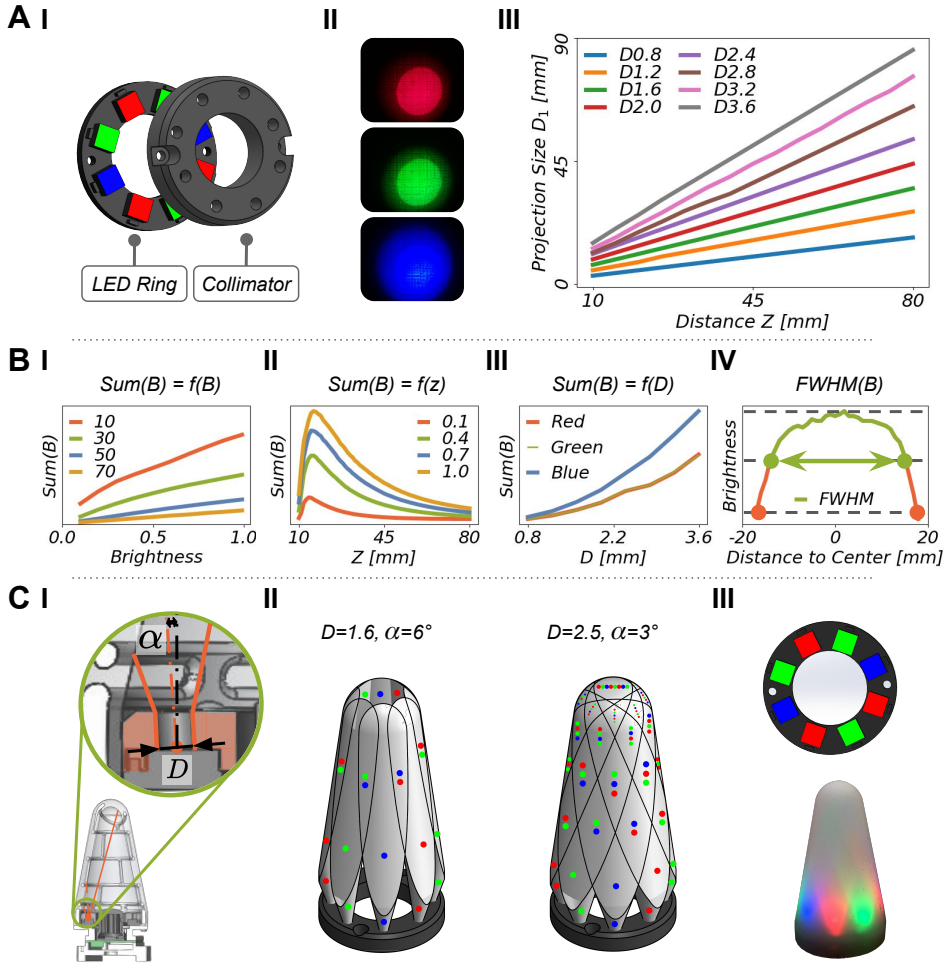

Figure 6.4: **Analysis of the lighting system**. **A** shows the correlation between collimator diameter D and light cone size. **A-I** and **A-II** depict the LED ring, the collimator, and light patterns projected on a flat plane from different red, green, and blue light sources. **A-III** shows the linear scaling of the cone size. Z is the imaging distance, D_1 is the diameter of the projection pattern, and different lines (D) are for different collimator diameters. **B** summarizes the light attenuation behavior. **B-I** shows the sum of light as seen by the camera depending on the surface distances (Z) for different brightness values of a single light source. **B-II** shows the same quantity as a function of the imaging distance for different brightness values. **B-III** shows the camera sensitivity to light brightness of different colors (red, green, blue) with varying D. **B-IV** is the light attenuation curve for a single bright disc (as in **A-II**). We use full width at half maximum (FWHM) to the quantify the size of the disc. **C** illustrates the effect of collimator hole size D and angle α on the light cones and the overall light pattern. **C-I** depicts the details of the collimator hole geometry. **C-II** shows the light covering the surface area with different D and α. **C-III** shows the light color arrangement (R, G, B, R, G, R, B, G) and visibility in a translucent shell.

Table 6.3: Mechanical properties of different sensing surface material candidates.

Material	ρ [g/cm^3]	V [cm^3]	m [g]	E [kPa]	ν	A_e [%]	Curing Time	Degas
EcoFlex 0030	**1.07**	**18.6**	**23**	**70**	**0.49**	**900**	**5 h**	**Yes**
EcoFlex 0035	1.07	18.6	23	70	0.49	900	10 min	No
EcoFlex 0050	1.07	18.6	23	80	0.49	980	4 h	Yes

Moreover, we find that the selected camera has different sensitivities to differently colored light sources: we observe a sensitivity ratio of 1 : 1 : 2 for red, green, and blue, respectively (Fig. 6.4B-III). Based on this sensitivity analysis, we arranged the light sources as R, G, B, R, G, R, B, G in series, and we simulate the light projection cores by tuning the collimator diameters and tilt angles, as seen in Fig. 6.4C.

6.3.2 Mechanical components

Insight's mechanical properties are optimized to ensure high sensitivity to contact forces, robustness against high impact forces, and low fatigue effects. For the elastomer, we choose EcoFlex 00-30 with a high maximum elongation ratio (900%) for sensing contact (Table 6.3). The skeleton is made of AlSi10Mg-0403 aluminum alloy, which can withstand forces up to 40 N in the shape of our prototype structure (Table 6.4). These two materials are chosen based on their material data sheets and finite element analysis (FEA) results [136]. All three of the mold pieces needed to cast the elastomer, as well as the skeleton, are 3D printed. We combine the skeleton and the elastomer without adhesive by over-molding, as described in Fig. 6.1B and Fig. 6.2B. The diameter of the skeleton beams and the thickness of the surrounding elastomer are optimized for robustness, as described in Fig. 6.5. FEA revealed that we can improve the system's sensitivity to contact forces by positioning the skeleton not in the center of the elastomer layer but closer to the inner surface.

Material for the soft shell We choose materials from the SmoothOn EcoFlex series for the sensor elastomer due to their wide application in soft robotics and their favorable properties in terms of weight, durability, and elongation ratio. We compare three materials out of this series in Fig. 6.5A and Table 6.3. We choose to use EcoFlex 00-30 due to its Young's Modulus, density, and curing time, as well as the fact that it can withstand the de-gassing procedure.

Material for the skeleton Considering weight, robustness, and yield strength, we design a beam structure as shown in Fig. 6.5B and compare different materials for the skeleton in Table 6.4. These three materials (stainless steel, aluminum alloy, and resin tough) can all be 3D printed. In the end, we choose the aluminum alloy due to its low weight. It can withstand up to 40 N without exceeding the yield strength limit, as suggested

Table 6.4: Mechanical properties of different sensor skeleton material candidates.

Material	ρ [g/cm^3]	V [cm^3]	m [g]	E [GPa]	ν	A_e [%]	σ_y [Mpa]	F [N]	σ [Mpa]	d [mm]
Steel	7.86	3.0	23.6	147	0.30	2.3	455	90	433.1	0.48
Aluminum	**2.68**	**3.0**	**8.0**	**73**	**0.33**	**4.1**	**227**	**40**	**192.4**	**0.46**
Tough	1.20	3.0	3.6	1.6	(0.49)	24.0	60.6	9	50.44	5.46

by a finite element model in which we apply a single nodal force at each meshed node of the skeleton structure.

Over-molding A robust connection is required between the elastomer and the skeleton over which it is molded. We test the minimal over-molding thickness of the elastomer as shown in Fig. 6.5C. We check whether the elastomer can cover a test cylinder without any defects. We find a minimal thickness for a robust connection to be around 0.8 mm.

Relative positioning The relative position of the skeleton inside the elastomer additionally affects the system's sensitivity to contact. We build a finite element model based on the material properties in Table 6.3 and Table 6.4 to analyze various possible relative positions between the skeleton and the elastomer. The analysis shows that positioning the skeleton with an offset near the internal surface will increase the sensitivity because this relative positioning causes more displacement – see Fig. 6.5D. In our design, the soft shell is 4 mm thick, and the skeleton is located 0.8 mm from the internal elastomer surface and 1.6 mm from the outer surface.

6.3.3 Optical properties

We need a material with the right reflective properties (surface albedo, specularity) for the sensing surface. On the one hand, it should not be too reflective because reflections saturate the camera and diminish sensitivity. On the other hand, no point on the surface should be very dark, because the camera needs to detect changes in reflected light. Moreover, the material has to prevent ambient light from perturbing the image and deteriorating the sensing quality. The sensing surface is made of a flexible and moldable translucent elastomer mixed with aluminum powder and aluminum flakes. The aluminum powder makes the surface opaque to ambient light, and the aluminum flakes adjust the reflective properties, as shown in Fig. 6.1C, Fig. 6.2B, Fig. 6.6 and Table 6.5.

We adjust the imaging object, namely the internal surface of *Insight*, based on the analysis of the light sources and camera settings. We test the material compositions and coloring for the sensing surface by mixing aluminum powder, aluminum flakes, and pigments into the elastomer. The goal is to obtain optimal reflective properties such that no part of the surface appears too dark or too bright, as shown in Fig. 6.6 and Table 6.5. In Fig. 6.6A-I, we investigate the effect of de-gassing (#0 using de-gassing

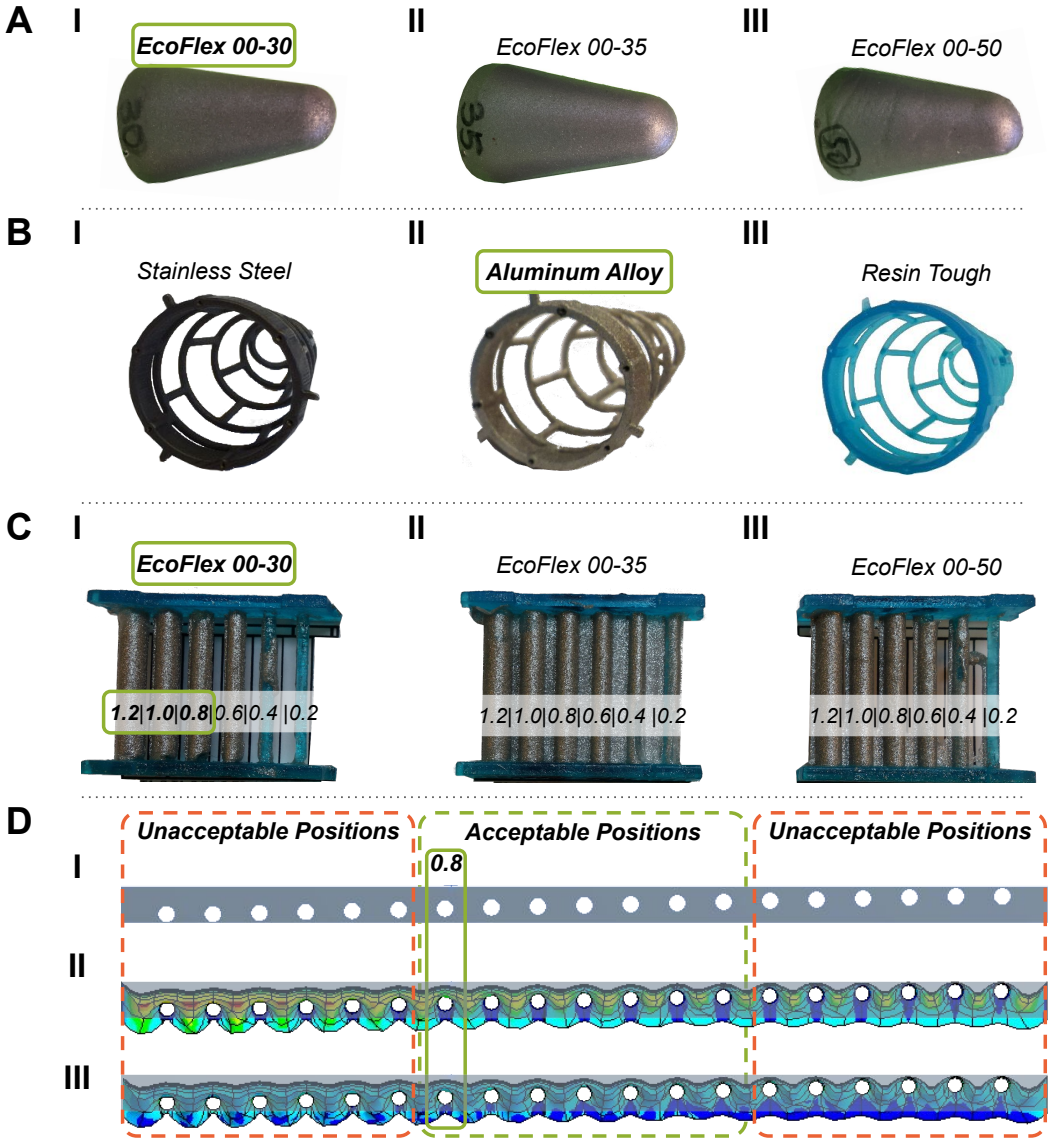

Figure 6.5: **Mechanical aspects of the sensor design**. **A** shows the candidate materials for the soft sensing shell: we choose EcoFlex 00-30. **B** shows the candidate materials for the stiff skeleton: we choose aluminum alloy. **C** presents the thickness test for the over-molding technique, which is used to connect the elastomer and the skeleton without adhesive. The minimal thickness for a robust connection is 0.8 mm. **D** shows a finite element analysis on how the relative position of the skeleton in the elastomer will affect the sensitivity, i.e., how much deformation occurs. **D-I** is a soft plate containing stiff rods at varying distances to the upper and lower surfaces. The left and right edges and the rods are fixed. Homogeneous pressure is applied to the upper surface. **D-II** shows the resulting deformation and **D-III** the induced von Mises stress.

Table 6.5: Different materials for sensing surface coating. The ten different material compositions considered for the soft sensing shell. The numbers are in given in grams [g] and correspond to the amount we used for molding one shell. Only about 25 g of the mixture is actually needed to fill the mold.

Material	0	1	2	3	4	5	6	**7**	8	9
EcoFlex0030-A	20	20	20	20	20	20	20	**20**	20	20
EcoFlex0030-B	20	20	20	20	20	20	20	**20**	20	20
Aluminum Powder 65 μm	0	0	0	3	0	2.0	2.0	**2.0**	2.0	2.0
Aluminum Flake 75 μm	0	0	0	0	1.5	0.1	0.2	**0.3**	0.5	1.0
Black Pigment	0	0	0.8	0	0	0	0	**0**	0	0
Vacuum chamber (degas)	–	✓	✓	✓	✓	✓	✓	✓	✓	✓

vs. #1 without) during elastomer molding and find that de-gassing gives a clearer and stronger elastomer. We compare the effect of three additives on the resulting surface's albedo: black pigment (#2), aluminum powder (#3), and aluminum flakes (#4). We find that black pigment (#2) absorbs almost all light, so the camera can hardly see anything. The aluminum powder (#3) shows adequate performance but gives relatively dark images. In comparison, aluminum flakes (#4) tends to create saturated points very easily because of strong specularity. Thus we trade-off these effects and mix aluminum powder and aluminum flake with different ratios in the elastomer, as shown in Fig. 6.6A-II. We tested the light intensities for different elements on the LED ring and choose composition #7 for *Insight*. Figure 6.6B shows the red, green, and blue channels for the image captured for the whole imaging system. The red and green channels separate quite well, while the blue channel contains a bit of both red and green channel values. This channel mixing might be due to the camera's white balance function (although it was switched off) and the elastomer's material properties.

6.4 Results

The performance of the sensor is evaluated with respect to both accuracy and sensitivity. The first measure of accuracy is direct single-contact estimation: a contact force needs to be localized, and its magnitude and direction must be inferred. Second is force distribution estimation for single contact: the contact area and directional force distribution over the entire sensing surface are inferred. In addition, we provide qualitative results for multiple contacts. Lastly, we evaluate *Insight*'s sensitivity by providing statistics for small interaction forces, studying whether gravitational effects are perceivable, and characterizing its ability to detect shapes contacting the high-sensitivity zone.

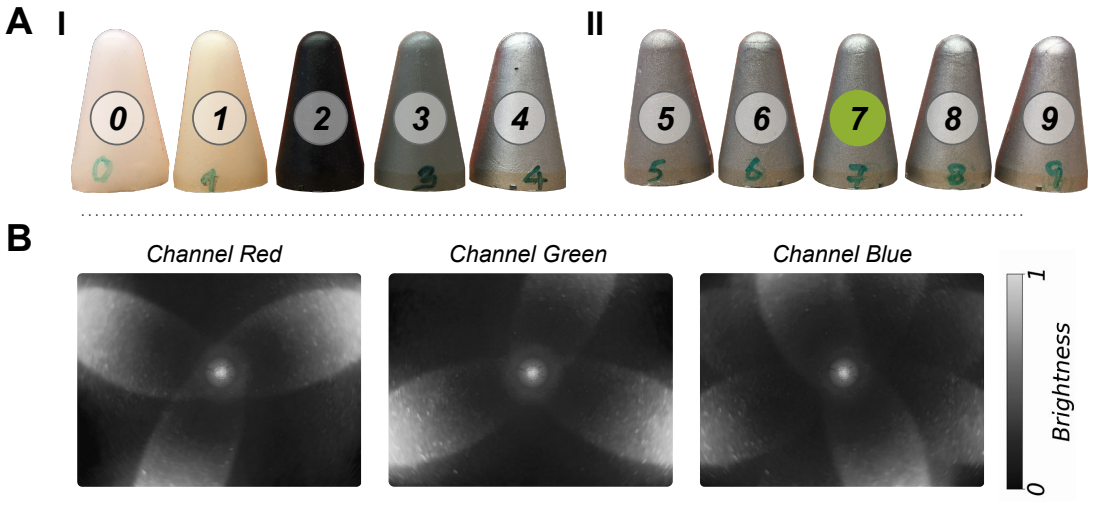

Figure 6.6: **Soft material composition**. **A** shows the ten tested versions of the EcoFlex material containing different pigments and additives, as defined in Table 6.5. **A-I** shows samples with no additive or only a single additive, while **A-II** presents a range of blends between aluminum powder and aluminum flake. **B** shows the red, green, blue channels of the light projection pattern formed by the chosen light sources shining on the inside of the selected soft shell (version 7).

6.4.1 Direct Contact Estimation

One of the most important ways to measure haptic sensor accuracy is to quantitatively evaluate the system's ability to localize contacts and measure the applied force. We employ a machine-learning-driven pipeline. First, we use a hemispherically tipped indenter with a diameter of 4 mm to probe a large number of points distributed across *Insight*'s sensing surface (Fig. 6.7A-I). In this procedure, we collect the images under contact and the contact force vectors from the ATI Mini40 force sensor, as well as the position of each contact on the sensor's surface using our 5-DoF test bed (Fig. 6.1D). The histogram of the applied forces in Fig. 6.7A-III shows that most contacts have magnitudes smaller than 1.6 N, as we set this value as the threshold of data collection to avoid damaging the sensor. Then we train a machine-learning model (modified ResNet [71] structure) to infer the contact information. The inputs to the model are the image under contact, a static reference image without contact, and a static image of the stiff skeleton for inhomogeneous elasticity encoding (recorded before over-molding in a dark environment). The outputs are the 3D coordinates of the contact in the sensor's reference frame and the 3D force components expressed in the local surface coordinate frame, as depicted in Fig. 6.7A-II. Details about data collection and machine learning are summarized in Section 6.4.6.

We evaluate the single-contact direct estimation accuracy of localization and force

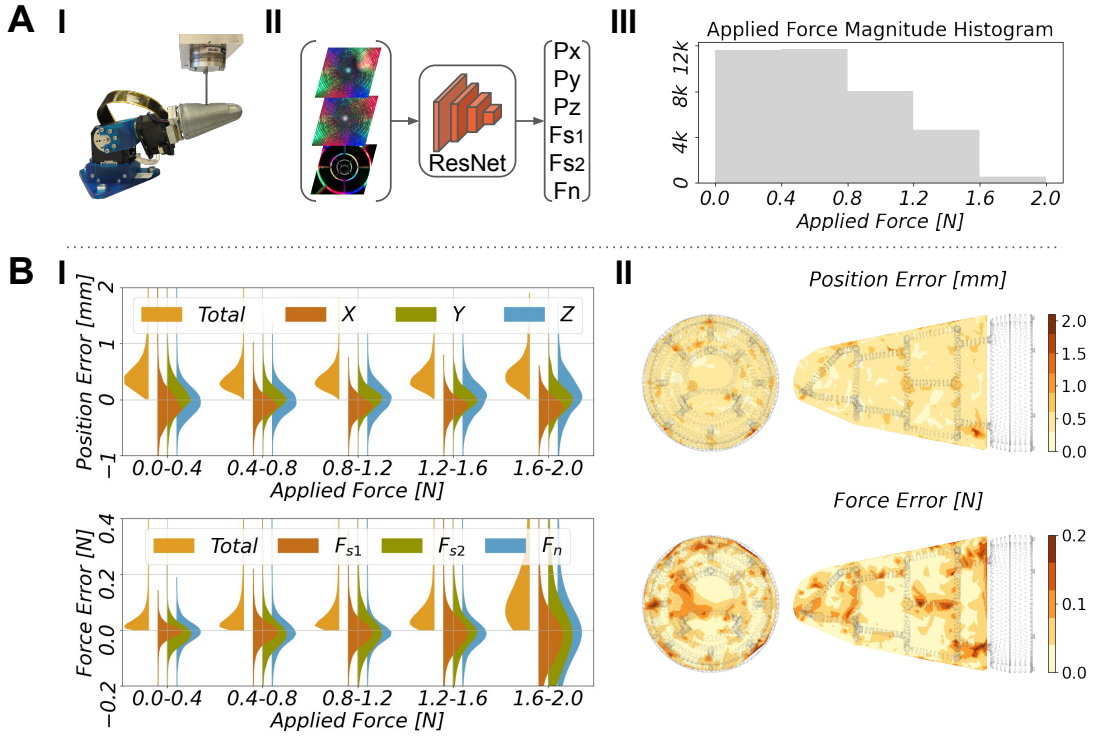

Figure 6.7: **Single contact performance with direct estimation**. **A**: the estimation pipeline for inferring single contact position and force. **A-I** shows the real experimental setup, in which the test bed probes *Insight* and collects data. **A-II** sketches the machine-learning model: the inputs are three images (raw image, reference image, and skeleton image), and the outputs are the contact location and contact force vector. **A-III** shows a histogram of the forces applied in the data collection procedure. **B**: statistical evaluation of the sensor's performance on the test data. **B-I** presents the localization and force estimation performance grouped by applied force magnitude. The red-, green-, and blue-colored half-violins show the distribution of deviations in the x, y, and z directions, respectively. The force is predicted relative to the surface in normal direction F_n and two shear directions F_{s1} and F_{s2}. The orange half-violins stand for the resulting total errors. **B-II** indicates the spatial distribution of the localization and force quantification errors for the same test data.

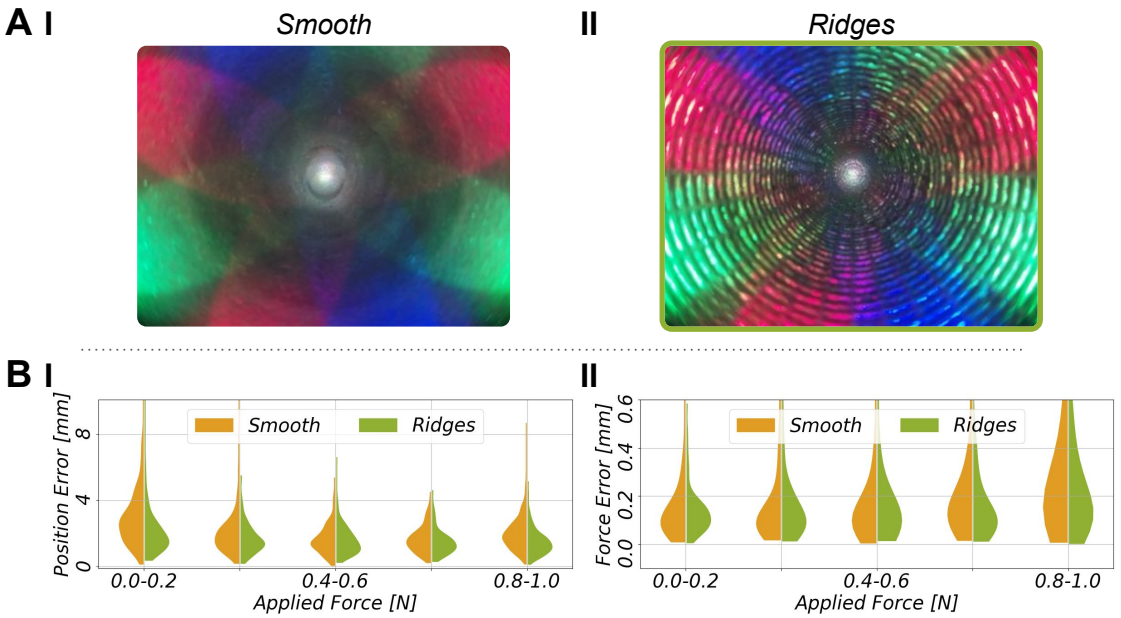

Figure 6.8: **Effect of the surface structure on the sensor performance**. **A** indicates two versions of internal surface morphology: one is a smooth surface, and another is with ridges. **B** compares the sensor performance of localizing single contact (**B-I**) and force quantification (**B-II**) with different force strengths, respectively. The errors in the case of the ridged surface (right half-violins) are generally lower than for the smooth surface (left half-violins).

sensing for an applied force magnitude up to 2.0 N, as shown in Fig. 6.7B-I. All reported numbers are for test contact points that do not appear in the training data. The overall median localization precision is around 0.4 mm, and the force magnitude precision is approximately 0.03 N in the normal and shear directions. The force direction is estimated with a precision of approximately 5°. Notice that the test bed has an overall position precision of 0.2 mm, and the force-torque sensor has a force precision of 0.01/0.01/0.02 N ($F_x/F_y/F_z$). *Insight*'s accuracy in localization is remarkably stable over different force ranges, while the error in force amplitude slightly increases with higher interaction force. For strong applied forces (over 1.6 N), the force accuracy becomes worse, presumably because we have little training data for this domain (histogram in Fig. 6.7A-III). Another explanation is that high forces occur most often at locations near the stiff frame (Fig. 6.7B-II), which deforms only a little. There is no noticeable difference in the localization and force accuracy in the sensor frame's x, y, and z directions.

We particularly evaluate *Insight*'s accuracy at localizing test contact points, as shown in Fig. 6.7B-II. The accuracy is stable across the entire surface, and higher errors appear near the stiff frame. Only areas near the camera show a systematic performance drop. Because our camera has a 4:3 aspect ratio, it cannot see two opposite areas at the base of

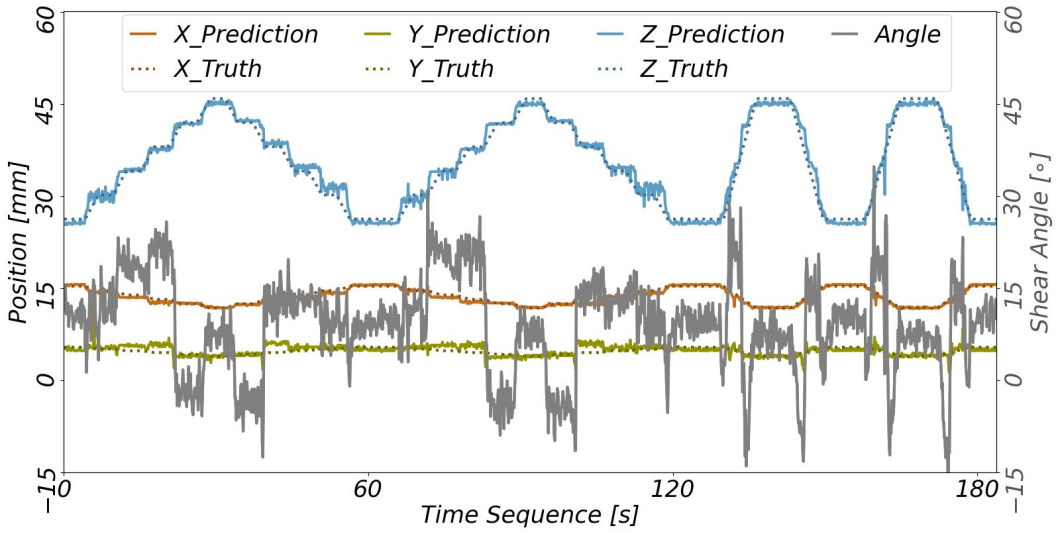

Figure 6.9: **Localizing the indenter in sliding motion**. This plot shows the sensor's capability of localizing an indenter when it is sliding along the sensor surface. The dashed lines show the actual position of the indenter over time, and the solid lines are the estimated contact locations. The listed axes refer to the global coordinate frame of *Insight*. The gray line shows how the force angle changes during the sliding motion. It is the angle between the estimated force vector and the sensing surface normal vector.

the shell, below the lowest ring of the stiff frame.

Surface morphology We investigate the effect of the surface morphology on the sensor performance. As shown in Fig. 6.8, we compare a smooth inner surface to a surface with ridges. Taking the single contact evaluation performance as the criterion, we find the surface with ridges improves the localization and force quantification, as seen in Fig. 6.8B. Moreover, we empirically find that the surface with ridges helps to accelerate the machine learning training procedure. Furthermore, it is easier to track the movement of the ridges than that of the smooth surface; one should be able to design more advanced computer vision algorithms to improve sensor performance.

Dynamic tracking We quantitatively show the sensor's performance at localizing an indenter in sliding motion. The experimental setup is as follows: the indenter first contacts *Insight* and then slides along the sensor surface for 4 mm before stopping for 5 seconds. This sliding-and-stopping behavior is repeated five times in the forward direction and five times in the backward direction. We use two complete cycles of this behavior. After this, the indenter slides along the sensor in the forward direction and the backward direction without any pauses for another two cycles.

The contact can be accurately localized, and the direction of the indented force can be

discriminated as shown in Fig. 6.9. We evaluate the changes of the angle between the estimated force vector and the sensing surface normal vector during the sliding motion. The gray line in Fig. 6.9 shows the angle change at each beginning of sliding motion with a recovering phase during the pause interval. Between the sliding segments, there is one position that shows an abnormal angle change. It is caused by the metal beam of the skeleton.

6.4.2 Force Map Estimation

To infer contact areas and multiple simultaneous contacts, we now consider the distribution of contact force vectors across the entire surface, which we call a whole-surface force map. Altogether, the force map yields valuable information for robotic grasping and manipulation, e.g., for slip detection, in-hand object movement, and haptic object recognition.

The *Insight* sensor has a 3D curved surface and thus needs to output a force map with the same shape. We create a fine mesh of points spanning the entire surface with an average distance of 1 mm between neighboring points. For the results reported here, we use 3800 points. Each point has three output values describing the force components it feels in the x, y and z directions expressed in the reference frame of the sensor.

Similar to the direct contact estimation, we also develop a machine-learning-driven pipeline. Instead of the six-dimensional output (Fig. 6.7A-II), the network now produces the approximate force distribution map (Fig. 6.10A-I) using only convolutional layers. The map is estimated as a flat image with three channels ($\mathbf{F_x}, \mathbf{F_y}, \mathbf{F_z}$) to describe nodal forces (individual force on each point) in the x, y and z directions, respectively, mimicking the red, green, and blue channels in a colorful image. Each pixel in the image corresponds to one point in the force map. The correspondence is established using the Hungarian assignment method [183], which minimizes the overall distance between pixels and points projected to the 2D camera image, as shown in Fig. 6.10A-II. Training the machine-learning model from collected data additionally requires target force distribution maps (Fig. 6.10D). Since they are not measured directly, we approximate the force map applied by the indenter by distributing the measured total force locally across the surface. From a set of five diverse candidates, the approximation yielding the best performance in localization and force magnitude accuracy is selected (see Fig. 6.11, Table 6.6).

The quantitative estimation accuracy for the force amplitude and force direction are reported in Fig. 6.10B grouped by force magnitude. The evaluation is based on the comparison between the three-dimensional force vectors summed across the predicted force map and the ground-truth force vectors using the same single-contact data set. The median error in inferring the total force is around 0.08 N, and the error grows with increasing force (Fig. 6.10B-I and Fig. 6.12). The system's tendency to slightly underestimate larger forces is likely caused by our force map approximation method, the influence of the skeleton, and the machine-learning method itself, which tends to estimate smooth force distributions rather than peaked maps. The median error in inferring the

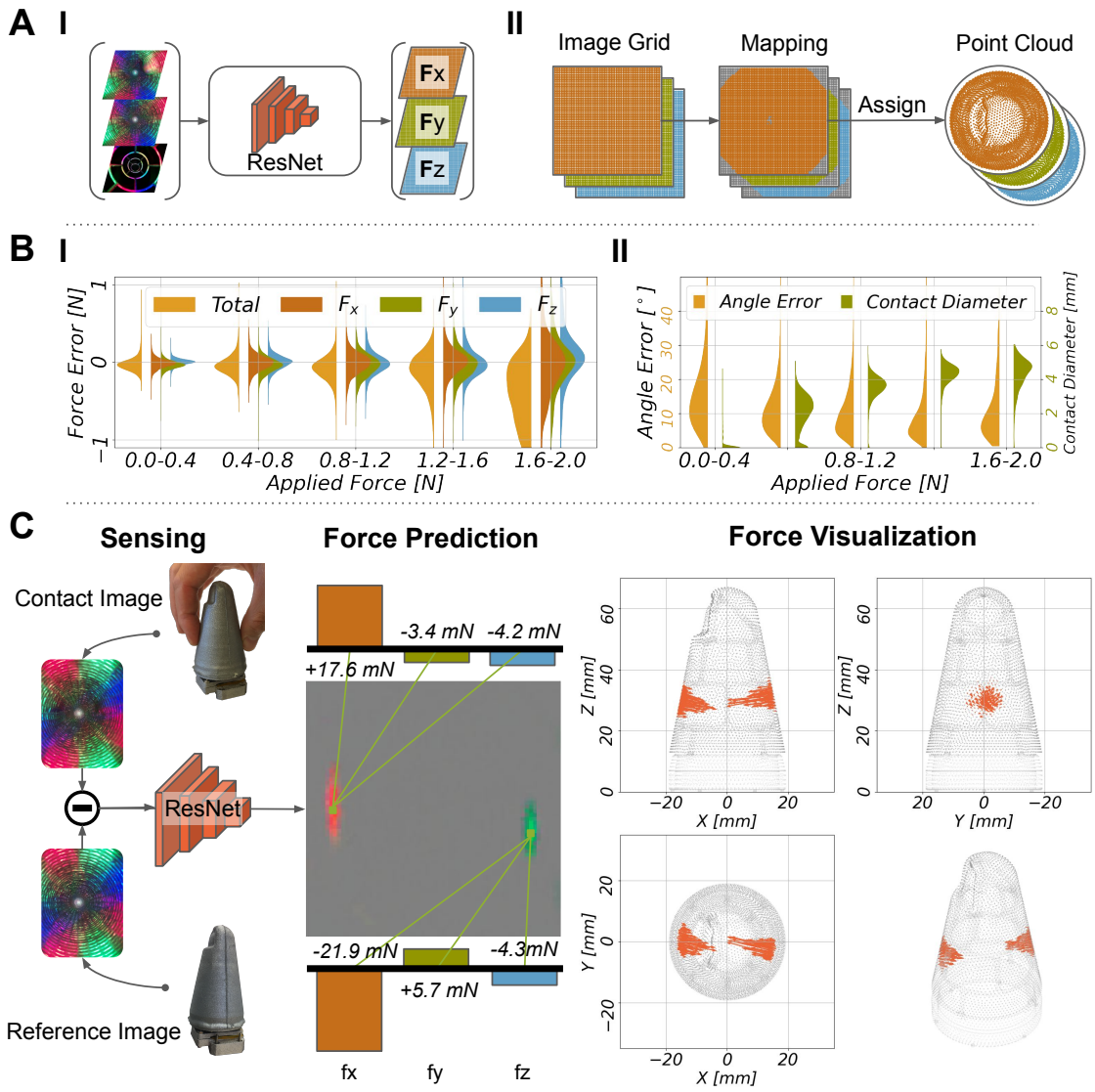

Figure 6.10: **Performance evaluation of the force map.** **A** indicates the pipeline of estimating the force distribution; the ResNet network transforms three images (raw, reference, skeleton) into the x-y-z force map image (**A-I**), and its pixels are mapped to points on the sensing surface (**A-II**). **B** shows the quantitative evaluation of the performance for force amplitude, force direction, and contact area size inference grouped by applied force amplitude. **C** demonstrates the data flow and estimated force map when the sensor is pinched and rotated by two fingers.

force direction is around 10° for low contact forces, and it decreases to 5° with higher applied forces (Fig. 6.10B-II). Moreover, we can also localize the contact with a precision around 0.6 mm based on the force map by averaging the locations of the 20 points with the highest force amplitudes (Fig. 6.12). The contact area is estimated by identifying the points with predicted forces larger than 0.02 N. The diameter of this contact area increases with higher applied force and tends to overestimate by about 1 mm for a 4-mm indenter at high forces.

Indenter approximation To generate the force map data set, we need a force approximation model to describe the force distribution when the indenter is contacting *Insight*'s outer surface. We measure the total applied force vector with the force-torque sensor at the indenter (Fig. 6.1D part 2). One theoretical approach to modeling how this force is distributed is to use Hertz contact theory [184]. For a spherical indenter with radius $\sigma = 2$ mm contacting an elastic half-space through a contact area with the same radius, this distribution is given by the following formula:

$$F_H(x) = F\frac{1}{Z}\begin{cases}\left(1-\frac{x^2}{\sigma^2}\right)^{\frac{1}{2}}, & \text{if } |x| \le \sigma \\ 0, & \text{otherwise}\end{cases} \tag{6.1}$$

where x is the radial distance from the contact center point, F is the measured force, and Z is the normalization constant such that the integral of the profile is 1.

However, Hertz theory is not appropriate for thin elastic sheets that deform as a whole under the force of the indenter. More importantly, the Hertz profile causes problems in our machine-learning procedure because it strongly localizes the target signal. We have a mesh grid of the sensor's outer surface, and this grid has 3 800 points with neighboring points separated by around 1 mm. The indenter has a diameter of 4 mm. Using the Hertz profile causes at most 13 points among 3 800 points to have a non-zero value, which we found empirically to be hard to train using the machine-learning procedure. Thus, we check four alternative profiles to understand this issue and verify the benefits of distributing the force more widely, as illustrated in Fig. 6.11. First we consider a Laplacian profile with a cutoff at 2σ:

$$F_L(x) = F\frac{1}{Z}\begin{cases}e^{-\frac{|x|}{\lambda}}, & \text{if } |x| \le 2\sigma \\ 0, & \text{otherwise}\end{cases} \tag{6.2}$$

and two shapes: Laplacian1 with $\lambda = 0.87\sigma$, which has the same maximal value as the Hertz model, and Laplacian2 with $\lambda = 0.5\sigma$, which is more peaked. Another alternative is a truncated Gaussian distribution:

$$F_G(x) = F\frac{1}{Z}\begin{cases}e^{\left(-\frac{1}{2}\cdot\frac{x^2}{0.4\sigma^2}\right)}, & \text{if } |x| \le 2\sigma \\ 0, & \text{otherwise.}\end{cases} \tag{6.3}$$

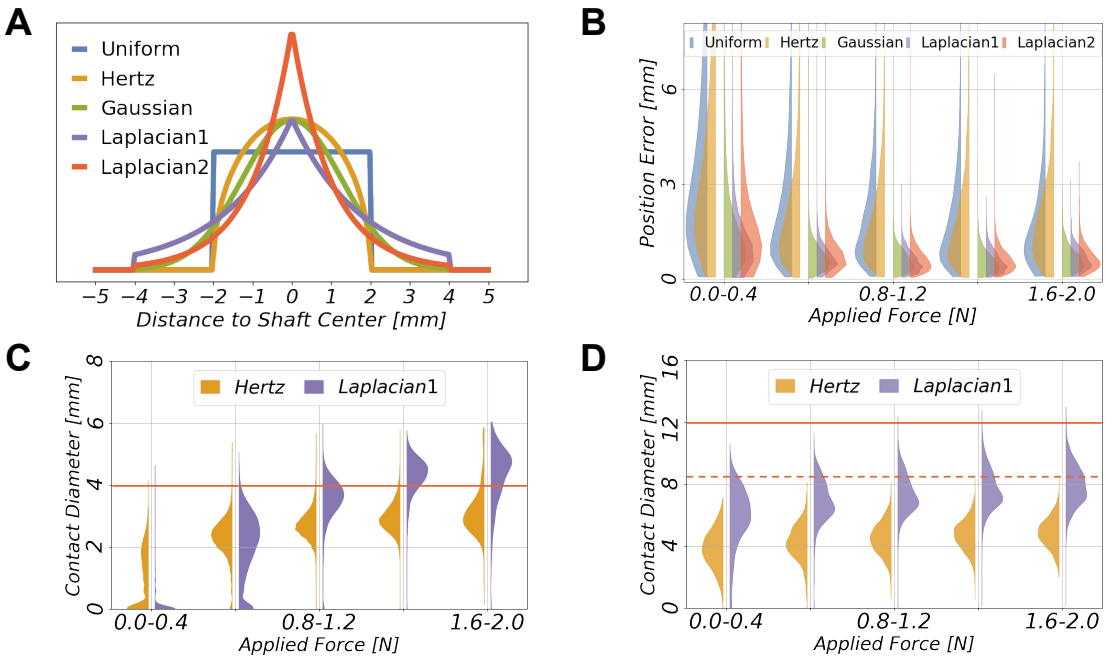

Figure 6.11: **Force distribution approximation and evaluation**. **A**: The five approximate force distribution methods that we tested for the 4 mm sphere-shaped indenter. The illustration is in one dimension and was revolved to distribute the measured contact force vector across the local surface of the force map around the contact point. The uniform (blue) and Hertz theory (orange) distribution curves are strongly localized with a radius of 2 mm. The green curve follows a truncated Gaussian distribution, and the two other curves (purple and red) follow Laplacian distributions that also stop at a radius of 4 mm. **B** shows the position inference performance using the different methods. **C** and **D**: The diameters of the contact area prediction for two indenter sizes (4 mm, and 12 mm) using two different approximation methods (Hertz and Laplacian1). The dashed line indicates a reference diameter from Fig. 6.13A-I based on the fact that the large indenter penetrates into the sensor only partially.

Finally, we tested a uniform distribution with the indenter's radius as a reference:

$$F_U(x) = F\frac{1}{Z}\begin{cases}1, & \text{if } |x| \leq \sigma \\ 0, & \text{otherwise.}\end{cases} \tag{6.4}$$

A comparison of the different models and the resulting performance are shown in Fig. 6.11 and Table 6.6. We find that the approximation maps with a smoother profile (Laplacian and Gaussian) achieve better position accuracy than those with sharp edges. In particular, the Laplacian1 with its flatter profile yields the best accuracy in position,

Table 6.6: Evaluation of force distribution approximation. Evaluation on five different force distribution approximation curves for the 4 mm indenter. The first three presented values are median errors of different types. The two values in the "Contact Diameter" column show the median contact area size based on two force thresholds for counting contact points; the first diameter uses a force threshold of 0.02 N, and the second uses 0.01 N.

Shape	Position Error [mm]	Force Error [N]	Angle Error [°]	Contact Diameter [mm]
Uniform	1.9	0.10	11.1	2.4, 2.9
Hertz	1.8	0.11	12.4	2.4, 2.7
Gaussian	0.7	0.10	11.7	2.6, 3.7
Laplacian1	**0.6**	**0.08**	**10.2**	**2.4, 4.2**
Laplacian2	0.7	0.12	13.7	2.2, 3.5

force magnitude, and force direction prediction. All approximations tend to have similar contact area predictions, with a slight overestimation for the smoother profiles for strong forces. Based on these results, we choose *Laplacian1* (Eq. 6.2 with $\lambda = 0.87\sigma$) as our force-distribution approximation method.

Our interpretation of why the smoother profiles achieve better results is that they cover a larger neighborhood and therefore reduce the sparsity in the target signals. The machine-learning model has the tendency to produce a smoothed output, so the peak of our approximation model is smoothed out.

An alternative to the approximation would have been a finite-element method to compute the local force distribution. We did not pursue such an approach for two reasons. First, with our simulation tools the simulation of all tested contact locations would take around 50 days. Second, the linear assumption between deformation and force in the simulation is violated in our sensor design due to large deformations.

Force map evaluation Using the force map as the target output for our training data allows us to predict the force distribution directly from an image. The performance of the system is summarized in Fig. 6.10 in terms of force quantification and force direction estimation. Fig. 6.12 presents a more detailed evaluation of the performance. First, we want to quantify single-contact precision. We select the 20 points from the force map with the highest predicted force magnitudes and take the mean of their positions. The localization performance shows no visible differences in the x, y, or z directions; see Fig. 6.12 A & B-I. However, we get slightly worse results than our direct estimation approach (Fig. 6.7). Direct prediction has a median error of 0.4 mm and 0.03 N, whereas the inference computed from the force field has a median error of 0.6 mm and 0.08 N. Second, we consider the force magnitude, which is displayed in Fig. 6.12B-II. The estimated force tends to under-estimate the actual forces for strong applied forces. Third, we present *Insight*'s performance in estimating the force direction. Figure 6.12C shows

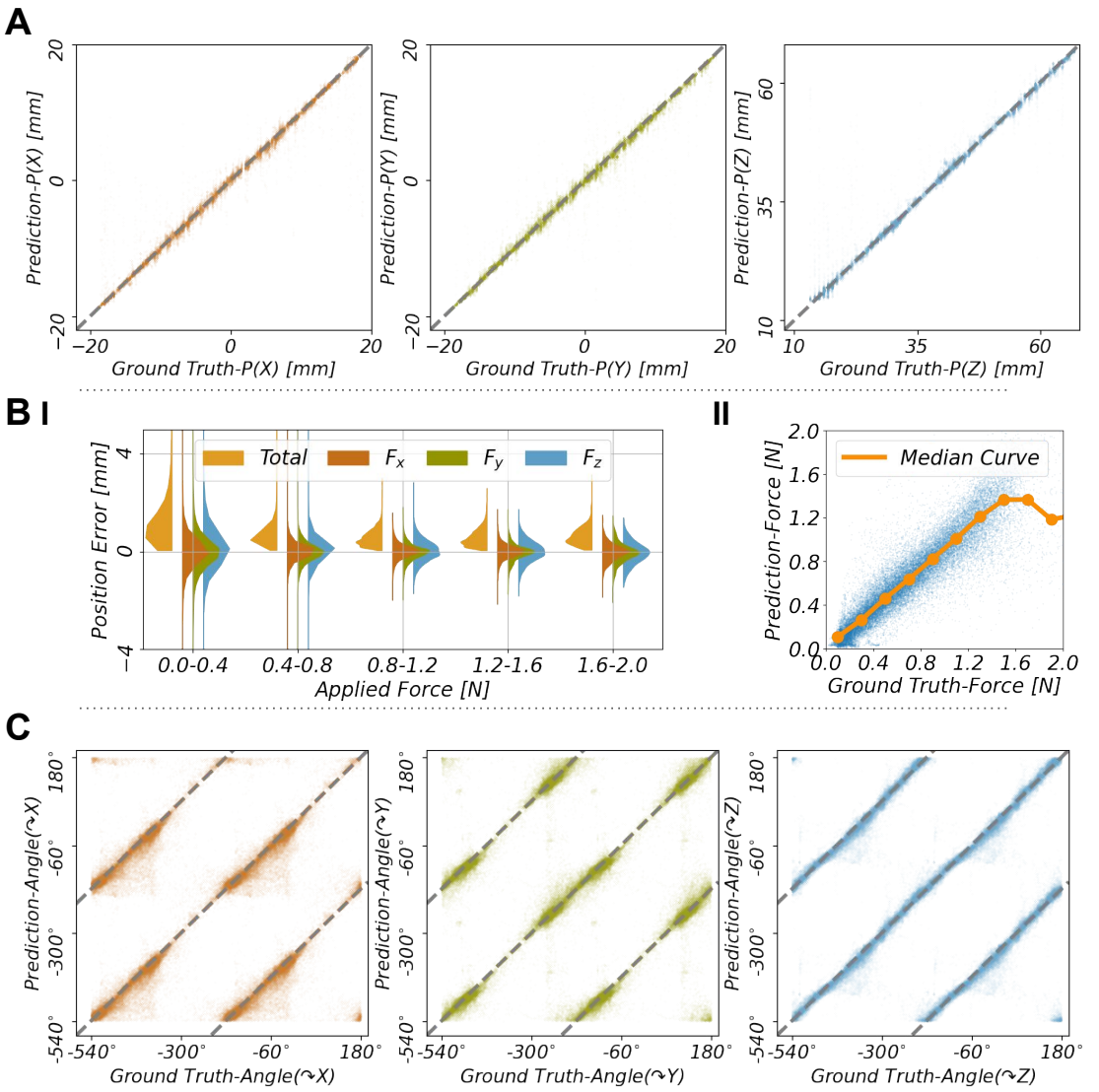

Figure 6.12: **Force map evaluation**. Based on the estimated force map, we extract information about contact position (**A**, **B-I**), force amplitude (**B-II**), and force direction (**C**). **A** shows the ground truth and estimated positions in x, y, z directions, where **B** contains the corresponding statistical evaluation. **B-I** presents the localization performance grouped by applied force strength. The red, green, and blue colored half-violins show the distribution of deviations in the x, y and z directions, respectively. The orange half-violins are the resulting total errors. **B-II** shows the estimated force magnitude as a function of the ground truth force. The median curve indicates a good overall correspondence with a tendency toward underestimation for larger forces, partially caused by a paucity of data in this regime. **C** presents the estimated force directions indicated by the angles around the x, y, and z axes. The gray dotted lines indicate perfect prediction. For better visualization, and to avoid cropping, the angle range is extended to ($-540°$ to $180°$).

the true and predicted force angles relative to the three coordinate axes x, y, and z. Overall, we observe a very good correspondence. The sensor can estimate the rotation direction around the z-axis better than the other two, while the x and y directions show similar performance.

Force map evaluation for a larger indenter We quantitatively validate the ability of our machine-learning pipeline to generalize the force map inference problem by conducting experiments with a larger indenter. The same test bed was used to probe *Insight* along its entire outer surface with a hemispherical indenter with a diameter of 12 mm, which is three times wider than the indenter with which the training data set was collected. We then process the data through the previously trained machine learning model and compare it with the ground truth.

The geometry of the indenters, the elastomer, and the stiff frame are visualized in Fig. 6.13A-I. The evaluation is on 16 919 indentation samples with widely varying applied force magnitudes, as shown in the histogram of Fig. 6.13A-II. The force threshold for the data collection is set to 3 N, but almost all collected data points have a total force less than 2 N. This pattern indicates that at some positions the large indenter experiences a sudden force increase, quickly raising the total resultant force over the threshold. This situation happens when the indenter hits the stiff frame directly. Given the lack of data at higher forces, we evaluate the sensor performance with an applied force threshold of 2 N.

Fig. 6.13B shows the results of this quantitative evaluation. The median position error is 1 mm; it is largest (1.6 mm) at smaller indentation forces, and it gets smaller (0.8 mm) with higher indentation forces. The median force error is 0.23 N; it is smallest (0.09 N) at smaller indentation forces, and it gets bigger (1 N) with higher indentation forces. The median direction error of the force is 16°; it is largest (16°) at smaller indentation forces and gets smaller (9°) with higher indentation forces. The median estimated contact area diameter increases from 6 mm to 8.5 mm as the contact force increases, which approximates Hertzian contact theory for a large indenter with relatively small indentations.

In addition, we include a qualitative evaluation on the force map prediction, as shown in Fig. 6.13C. When the indenter presses on the soft material far away from the stiff frame, the force map shows a symmetric force distribution. When the indenter contacts an area near the stiff frame, it shows an asymmetric force distribution. To be more explicit, the strongest estimated contact force vectors appear directly at the frame, and smaller force vectors occur nearby, as would be expected from an inhomogeneous surface like this.

6.4.3 Multiple simultaneous contacts

We also qualitatively demonstrate the sensor's performance during multiple complex contacts, as shown in Fig. 6.10C and Fig. 6.14. The figures show how the captured image and a reference frame without contact are combined to yield the system's perceptual

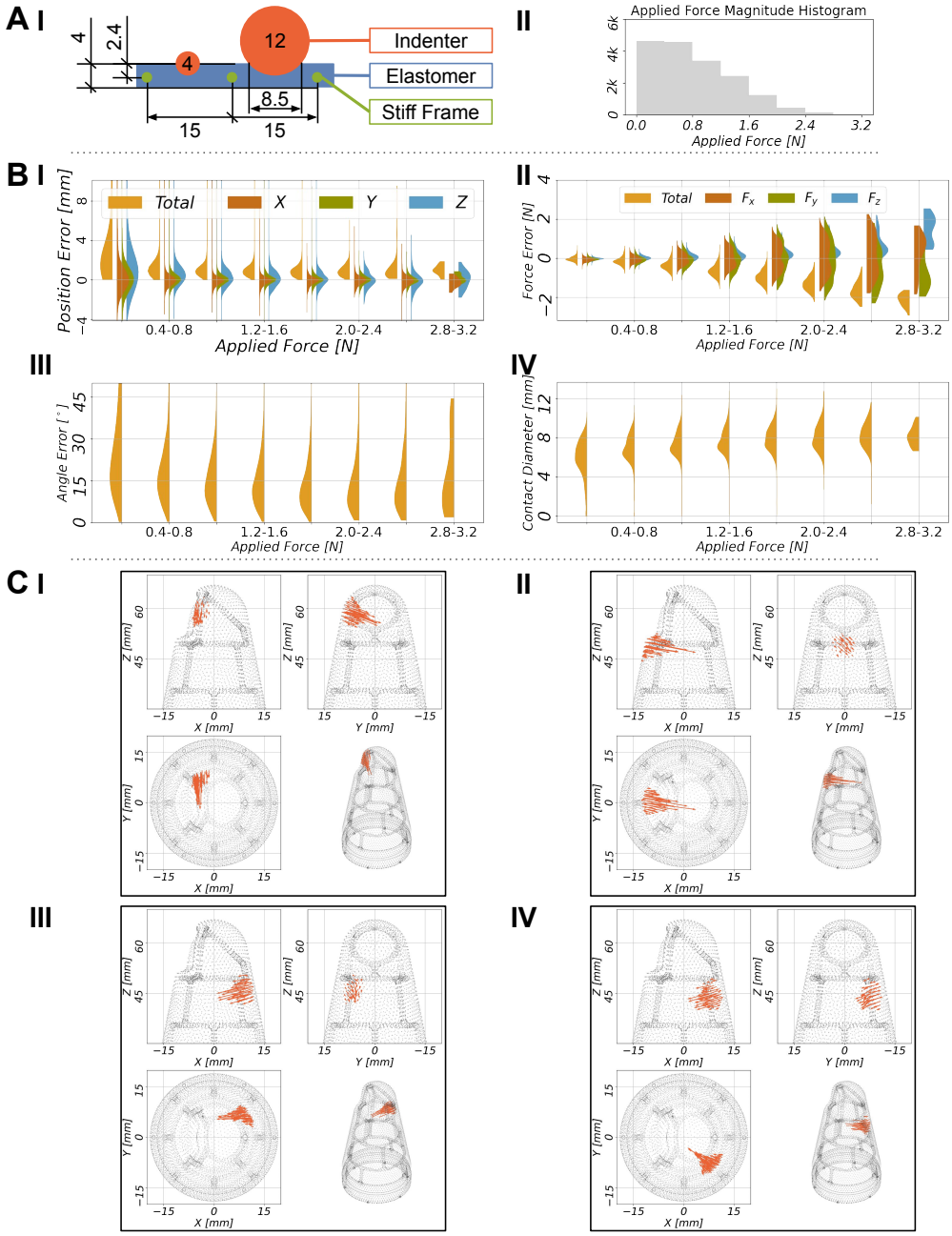

Figure 6.13: **Force map evaluation with a larger indenter**. An indenter with 12 mm diameter is used to validate the sensor performance. **A-I** shows the sizes and positions of the relevant components; **A-II** shows a histogram of the applied force. **B** presents an evaluation of test accuracy for localization, force strength, force direction, and contact area diameter. **C** demonstrates four sample cases of the force map prediction. C-I–III show an asymmetric force prediction that is plausible as the indenter is hitting a beam. C-IV shows the case without beam and the prediction is symmetric, as expected.

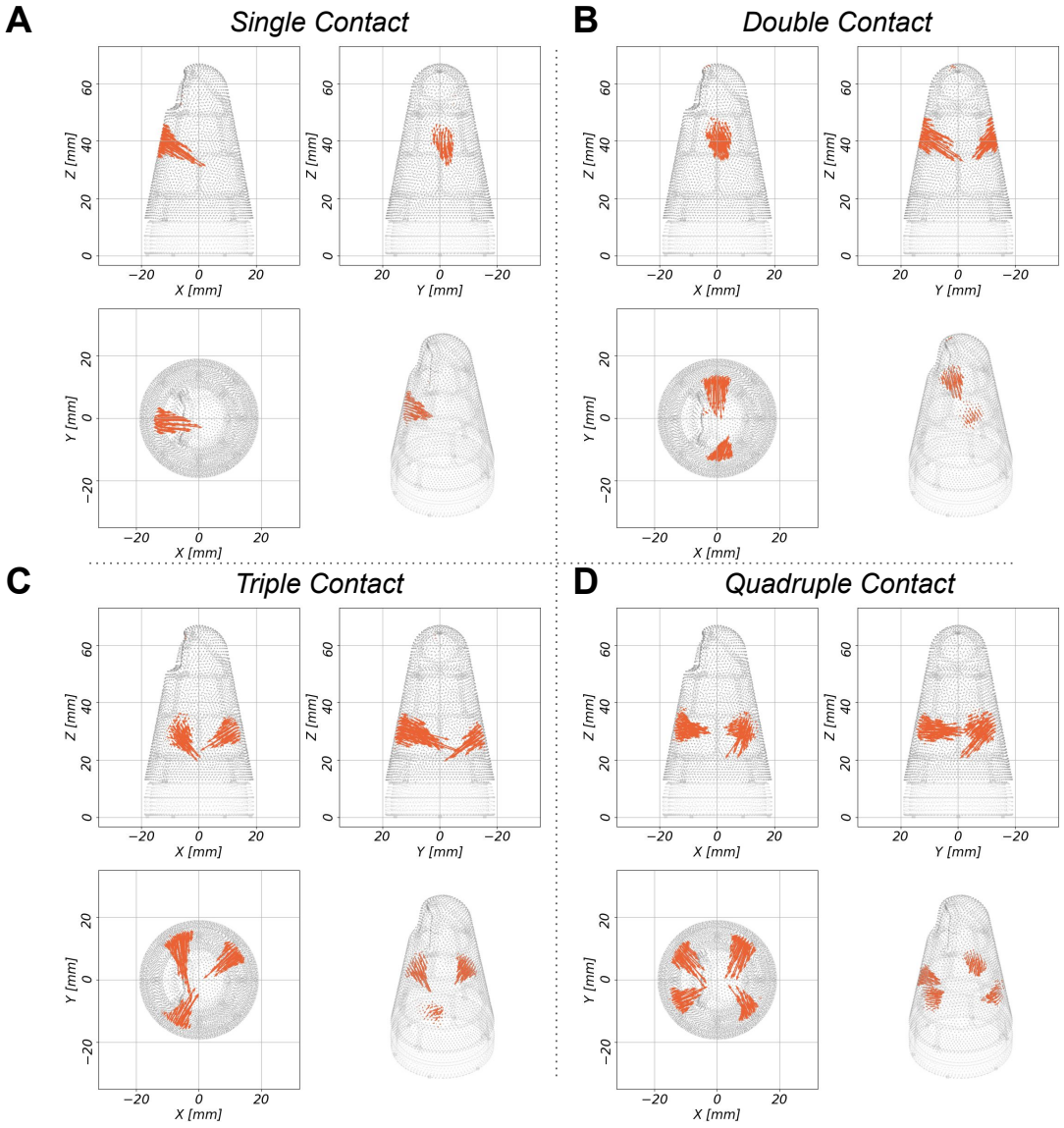

Figure 6.14: **Multiple contacts examples**. Visualizations of the force map distributions over the sensing surface for a single contact (**A**), double contact (**B**), triple contact (**C**), and quadruple contact (**D**).

response to a human using two fingers to pinch and slightly twist the sensor. Fig. 6.14 shows the response for many different contact situations. Each pixel of the force map contains the three force values estimated at that point. To facilitate interpretation, we also visualize each contact force vector on the 3D surface of the sensor. The experimenter's counter-clockwise twisting input can be seen in the slant of the force vectors when the sensor is viewed axially. In our experiments, the sensor was consistently able to discriminate up to five simultaneous contact points and estimate each contact area in a visually accurate manner.

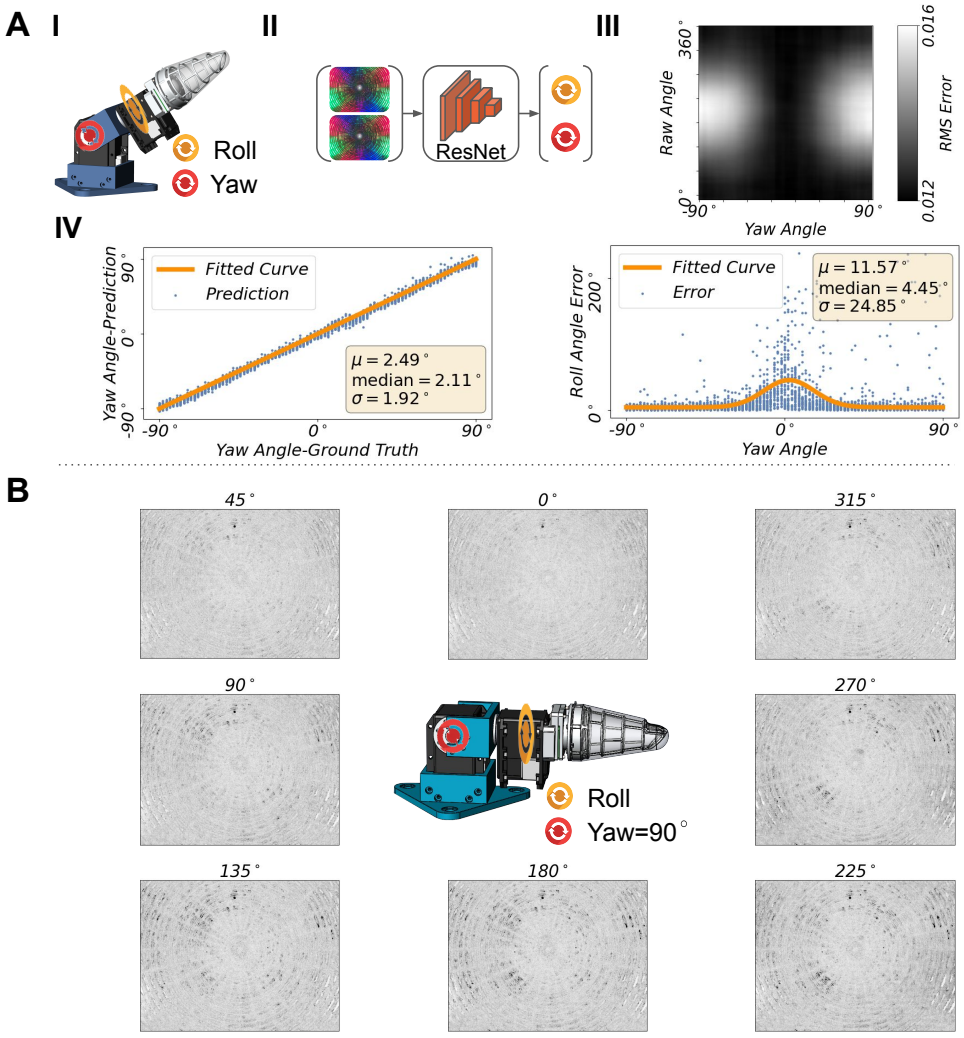

Figure 6.15: **Evaluation of *Insight*'s sensitivity to self-posture**. **A**: quantitative evaluation of sensor posture recognition. **A-I** and **A-II** present the experiment setup and the inference procedure for posture (roll and yaw angles). The network maps the difference between the current image and the reference image to posture coordinates. **A-III** shows the pixel-wise root-mean-square (RMS) of the image difference as the sensor is rotated to all possible roll and yaw angles. **A-IV** summarizes the posture estimation accuracy statistically: the yaw angle estimation performance with a yellow fitted curve (left) and the roll angle estimation accuracy under different yaw angles (right). **B**: Showcases for the deformation caused by gravity effect. It shows the image changes caused by gravity when the sensor rotates 360° around the roll direction while maintaining a yaw angle of 90°.

6.4.4 Sensor Posture Recognition

A haptic sensor created from soft materials also deforms due to gravity and inertial effects. In some contexts, these deformations will be considered disturbances that should be ignored in favor of contact signals. In other contexts, we can make use of these effects to enhance the capabilities of the sensor. Here, we demonstrate that gravity causes deformations that can be captured by *Insight*'s camera, and the sensor's posture can be recognized. To quantify the effect, we can train a machine-learning model to estimate the orientation of the sensor in terms of roll and yaw angles, as depicted in Fig. 6.15A-I & II.

In a quasi-static scenario, the sensor is held in the air with varying yaw angles (from -90° to 90°) and roll angles (from 0° to 360°). The raw difference between the captured images has a root-mean-square (RMS) difference per pixel of 0.012 due to noise and maximally 0.016 due to gravity, where 1 is the maximum possible (Fig. 6.15A-III). The estimation performance of the machine-learning model is provided in Fig. 6.15A-IV. Even from the slight deformation caused by gravity, the self-posture can be estimated with an overall accuracy of 2.5° in the yaw direction and 12° in the roll direction. The high error of estimating the roll angle appears when the roll axis aligns with the gravity vector, as would be expected. Figure 6.15B shows the image changes caused by gravity when the sensor rotates 360° around the roll direction while maintaining a yaw angle of 90°.

Despite the sensor's ability to estimate its own posture, the gravitational effect is still very small in practice and does not significantly affect the sensor's main functionality of perceiving external physical contacts, which are generally much larger than the self-weight of the elastomer skin.

6.4.5 Tactile Fovea

Our *Insight* prototype possesses a nail-shaped zone with a thinner elastomer layer, as indicated in Fig. 6.16A; with a sensing area of $13 \times 11\,\mathrm{mm}^2$, this tactile fovea is designed for detecting tiny forces and perceiving detailed object shapes. Based on FEA results and real experiments, we find that the thicker elastomer layer on the rest of the stiff skeleton smooths the shape of the contacted object, so that it is not easy to detect the exact shape of small objects. A very thin elastomer layer would be ideal but is also too fragile for vigorous interaction. We balance these two effects and choose a thickness of 1.2 mm for this special sensing zone.

The median position and force errors in the tactile fovea are 0.3 mm and 0.026 N over an applied force range of $0.03 - 0.8\,\mathrm{N}$, which shows better position accuracy and force accuracy than other sensing areas.

We conduct two demonstrations of how the fovea could be used for shape detection. The first one represents a v-shaped wedge with different levels of sharpness (included angles from 10° to 180°); see Fig. 6.16B and C. The second set of samples represents extruded polygons with an increasing number of edges (triangle, square, pentagon, hexagon, etc.); see Fig. 6.16D. Figure 6.16B shows that *Insight*'s camera can visually distinguish the

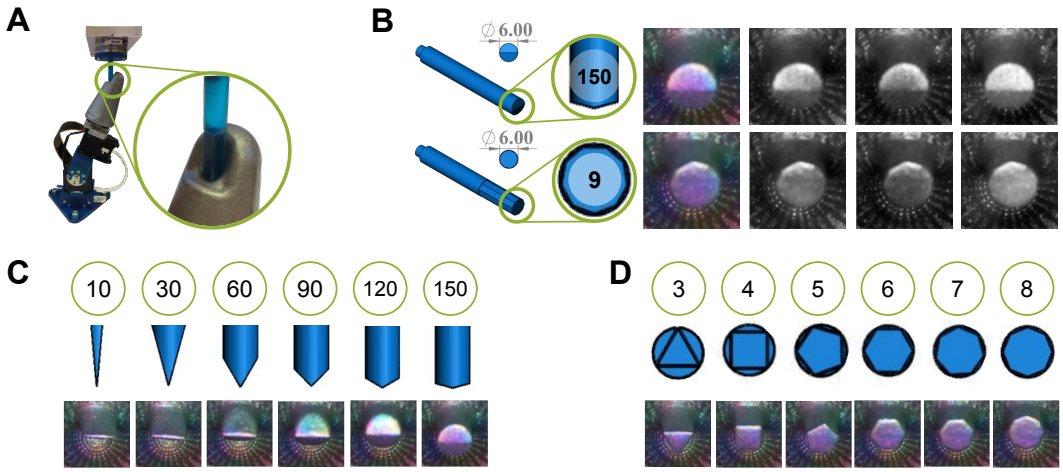

Figure 6.16: **Qualitative evaluation of the tactile fovea for shape detection**. **A** shows the experiment setup for applying differently shaped probes to the high-sensitivity region. **B** presents the system's perceptual limits for sharpness (v-shaped wedge with an included angle of 150°) and number of edges (nine-sided polygon) with an indenter diameter of 6 mm, along with their corresponding captured raw images and respective red, green, and blue channels. **C** and **D** depict the sharpness and edge tests: included angle and edge count (upper), indenter samples (middle), and captured images under indentation tests (lower).

tested samples up to 150° and 9-edges. An automatic procedure is left for future work. Theoretically, if the indentation depth increases, the shape detection accuracy will be improved. However, to not destroy the sensor by exceeding the elastomer's elongation capabilities, we limit the max indentation depth to 18 mm, which corresponds to half of the maximal deformation; the maximum resulting net force is 1.2 N in this area.

6.4.6 Machine Learning Details

Data Collection

Testbed We created a custom test bed with five degrees of freedom (DoF); three DoF control the Cartesian movement of the probe $(\vec{x},\vec{y},\vec{z})$ using linear guide rails (Barch Motion) with a precision of 0.05 mm, and two DoF set the orientation of the sensor (yaw, roll) using Dynamixel MX-64AT and MX-28AT servo motors with a precision of 0.09° and 0.2 mm. The probe is fabricated from an aluminum alloy and is rigidly attached to the Cartesian gantry via an ATI Mini40 force/torque sensor with a force precision of $0.01/0.01/0.02$ N $(F_x/F_y/F_z)$. The sensor is held at the desired orientation, and the indenter is used to contact it at the desired location.

Data Measurements are collected using our automated test bed to probe *Insight* in different locations. To obtain a variety of normal and shear forces, the indenter is moved to a specified location, touches the outer surface, deforms it increasingly by moving normal to the surface with fixed steps of 0.2 mm. For each such indentation level, the indenter also moves sideways to apply shear forces (normal/shear movement ratio 2:1). After a pause of 2 seconds to allow transients to dissipate, we simultaneously record the contact location, the indenter contact force vector from the test bed's force sensor, and the camera image from inside *Insight*. When the measured total force exceeds 1.6 N, the data collection procedure at this specified location terminates and restarts at another location. The contact location and measured force vector are combined to create the true force distribution map using the method described in Section 6.4.2. Images from *Insight* are captured using a Raspberry Pi 4 Model B with 2 GB RAM. All the data are collected and combined using a standard laptop.

Machine Learning

A ResNet [71] structure is used as our machine-learning model. The data for single contact includes a total of 187 358 samples at 3 800 randomly selected initial contact locations. The data set is split into training, validation, and test subsets with a ratio of $3:1:1$ according to the locations. The data for posture estimation from gravity contains 16 000 measurements and is split in the same way. We use four blocks of ResNet to estimate the contact position and amplitude directly (Fig. 6.7), two blocks to estimate the force distribution map (Fig. 6.10), and four blocks to estimate the sensor posture (Fig. 6.15). The machine-learning models are all trained with a batch size of 64 for 32 epochs, using Adam with a learning rate of 0.001 for mean squared loss minimization.

The raw images are interpreted into understandable haptic information using a custom machine-learning method. To tell when, where, and how the sensor is contacted, we present two formats to quantify sensor performances: direct single-contact inference with location and resultant directional (normal/shear) force, and three-dimensional force map over the 3D conical surface. In addition, the sensor posture is also inferred through a machine-learning method.

Our three information formats are trained using the same machine-learning-based architecture **ResNet-18** [71] but with customized modifications. The original ResNet-18 has one input convolution layer, one pooling layer to adjust the input size, and four standardized ResNet blocks afterwards connecting with a fully connected output layer. In all models we use an input image size of 410×308, which is down-sampled from the 1640×1232 image captured by the camera; the number of channels depends on the task.

Direct single-contact inference We use the standard ResNet-18 architecture with six input channels: three for the RGB difference between the input image and the reference image, and three for a static RGB image of the skeleton before overmolding (see Fig. 6.7). The fully connected output layer predicts six channels: the three positions (x, y, z) where

it estimates the contact is occurring, and the three force components (shear forces and normal force relative to the surface) at that location.

Force map inference For the force map inference, we modify the architecture. The input is the same as for the direct contact inference. However, we only use two ResNet blocks instead of four and replace the fully connected output by the following set of operations:

1. An appropriate up-scaling convolution using a transposed convolution to obtain the output dimension 64×64 from the 52×39 output of the second ResNet block
   ConvTranspose2d(128, 128, kernel_size=(7, 7), stride=(1, 1), dilation=(4, 2), output_padding=(1, 0)))
2. A convolution layer from 128 to 64 channels
   Conv2d(128, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))
3. Batch normalization followed by a LeakyReLU
4. A convolution layer to the force map (64×64 with three channels)
   Conv2d(64, 3, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))

There are several reasons for this architecture: the output is a spatial map where each output pixel is mostly influenced by a local region in the input. Thus, maintaining a good spatial resolution is advantageous. After two ResNet blocks, we still have a resolution of 52×39, which is similar to our desired output. Every additional block reduces the dimensions and adds considerable computation. The network has in total 1.5 million parameters. We are convinced that a smaller architecture can be found, but we leave this for future work.

Sensor posture inference For the task of predicting the posture of the sensor, we also use the standard ResNet-18 architecture. The input is the three channels of the difference image, and the outputs are the yaw and roll angles.

6.5 Discussion

We present a soft haptic sensor named *Insight* that uses vision and learning to output a directional force map over its entire thumb-shaped surface. The sensor has a localization accuracy of 0.4 mm, force magnitude accuracy of 0.03 N, and force direction accuracy of 5°. It can discriminate the locations, normal forces, and shear forces of multiple simultaneous contacts – up to five regions in our evaluation. Moreover, the sensor is so sensitive that its quasi-static orientation relative to gravity can be inferred with an accuracy around 2°. A particularly sensitive tactile fovea allows it to detect contact forces as low as 0.03 N.

How does *Insight* compare to other vision-based haptic sensors? Table 6.1 lists its performance along with that of twelve selected state-of-the-art sensors; we first give

an overview and then compare the designs. One of the earliest vision-based sensors is GelSight [18], which has a thin reflective coating on top of a transparent elastomer layer supported by a flat acrylic plate. Lighting parallel to the surface allows tiny deformations to be detected using photometric stereo techniques. Further developments of this approach increased its robustness (GelSlim [19]), achieved curved sensing surfaces with one camera (GelTip [107]) and with five cameras (OmniTact [108]), and included markers to obtain shear force information [109]. A different technique based on tracking of small beads inside a transparent elastomer is used by GelForce [49] and the Sferrazza and D'Andrea sensor [20, 175] to estimate normal and shear force maps. ChromaForce (not included in the table) uses subtractive color mixing to extract similar data from deformable optical markers in a transparent layer [110]. The TacTip [22] sensor family uses a hollow structure with a soft shell, and it detects deformations on the inside of that shell by visually tracking markers. Muscularis [112] and TacLink [111] extend this method to larger surfaces, such as robotic links, by using a pressurized chamber to maintain the shape of the outer shell; they are not listed in the table because they target a different application domain.

In terms of shape recognition and level of detail, the GelSight approach provides unparalleled performance. The tracking-based methods, such as GelForce and TacTip, are naturally limited by their marker density and thicker outer layer. *Insight* uses shading effects to achieve a much higher information density than is possible with markers, but its accuracy is also somewhat limited by measuring at the inside of a soft shell with non-negligible thickness. Beyond accurately sensing contacts, the robustness of haptic sensors is of prime importance. Without additional protection, GelSight-based sensors are comparably fragile due to their thin reflective outer coating, which can easily be damaged. Adding another layer increases robustness, but imaging artifacts were reported to appear after about 1500 contact trials because of wear effects [185]. We tested *Insight* for more than 400 000 interactions without noticeable damage or change in performance.

Each sensing technology imposes different restrictions on the surface geometry of the sensor. Vision-based tactile sensors need the measurement surface to be visible from the inside, so there is typically no space available for other items inside the sensor. The type of visual processing also matters. TacTip's need to track individual markers requires a more perpendicular view of the surface than shading-based approaches (GelSight and *Insight*). Soft materials deform well during gentle and moderate contact, but they do not withstand high forces if not adequately supported. GelSight uses a transparent rigid structure for support, which can lead to reflection artifacts when adapted to a curved sensing surface [107]. An alternative is high internal pressure [111], but then the observed deformations are non-local. The over-molded stiff skeleton in *Insight* maintains locality of deformations and withstands high forces.

To facilitate widespread adoption, tactile sensors need to be easy to produce from inexpensive components. Imaging components are remarkably cheap these days, making vision-based sensors competitive. However, GelSight needs a reproducible surface coating and permanent bonding between all layers, which are tricky to implement correctly [18, 186, 187]. TacTip needs well-placed markers or a multi-material surface that can be

3D-printed only by specialized machines. *Insight* uses one homogeneous elastomer that requires only a single-step molding procedure on top of the stiff 3D-printed skeleton. Being able to replace the sensing surface in a modular way increases system longevity; such replacement is supported by GelSight and TacTip in principle, and it is designed to be easy in *Insight*, though we did not evaluate the quality of the results that can be obtained without retraining.

Most sensors detect deformations with classical methods and use linear elastic theory to compute interaction forces. This approach requires good calibration and special care when it comes to reflection effects and inhomogeneous lighting. The linear relationship between deformations and forces is often violated for strong contacts and for inhomogeneous surfaces like the over-molded shell of *Insight*. Because our method is data-driven and uses end-to-end learning, all effects are modeled automatically. The downside of our approach is that it requires a precise test bed to collect reference data. Once constructed, the test bed can be used to collect data for different sensor geometries – only a geometric model of the design is required.

The inhomogeneity of our sensor's surface might cause unwanted effects in some applications. Robotic systems that move with high angular velocities and high accelerations will likely see tactile sensing artifacts caused by inertial deformations of the soft sensing surface; data collected during dynamic trajectories can potentially mitigate these effects. In addition, the neural network sometimes predicts forces around the base; these outputs occur because the sensor surface is deforming at the base connection points in a way similar to local contact. Furthermore, small areas of the surface near the base cannot be seen directly by the camera. Our future work will address this issue.

In general, our sensor design concept can be applied and extended to a wide variety of robot body parts with different shapes and precision requirements. The machine-learning architecture, training process, and inference process are all general and can be applied to differently shaped sensors or other sensor designs. We also provide ideas on how to adjust *Insight*'s design parameters for other applications, such as the field-of-view of the camera, the arrangement of the light sources, and the composition of the elastomer.

Chapter 7

Discussion and Outlook

Autonomous robots need adequate perceptions to perform real-world tasks. Robots can successfully perceive their surroundings using centrally mounted cameras and widely studied computer vision techniques. However, visual perception may fail in situations involving visual occlusion, particularly when robots contact other objects with unseen forces. As a result, robots may also need touch-sensitive skin to perceive visually occluded and small-scale deformation-involved contact information. Acquiring contact information often begins with the transduction of physical events to electrical measurements, which are subsequently calibrated to real-world physical values using data processing. In this thesis, we conceive and design four different haptic sensors to acquire contact information for applications with varied resolution requirements. This chapter will discuss the pros/cons of these four sensors, summarize our contributions to the haptic sensor community in Section 7.1, and discuss the potential challenges and research directions for machine-learning-driven haptic sensor designs in Section 7.2.

7.1 Contribution Summary

Problem A common theme of surface haptic sensors is to integrate many small sensing elements forming a grid along a flat or a curved surface. Each sensing element (taxel) is responsible for sensing interactions near its location locally. As done in [12], massive sensing elements are attached to a humanoid robot directly on the outer side. This design concept brings robustness problems. All the attached taxels directly interact with the environments, and all wires need to go through the whole surface to access and read out each taxel value. For varied applications, a resolution is desirable that would imply numerous taxels and wire connections. It is valid for small surface sensing, like fingertips, and large surfaces, around limbs. For fingertips, very focused areas need a high density of taxels with a fine size to perceive high-resolution haptic information, which is similar to touch screens. For limbs, even though the haptic information is coarsely needed referring to the density distribution of human mechanoreceptors, the sensing areas are very large and need many taxels. In general, a high density of smaller taxels and thinner wires are favorable to obtain a high resolution but less ideal due to mechanical fragility, electromagnetic noise, and crosstalk between taxels. Technical challenges also arise

concerning the physical size of the taxels and growing wiring and manufacturing costs for numerous taxels. I name the problem the dilemma of resolution and robustness.

Solution I propose to use the super-resolution concept to deal with this dilemma of resolution and robustness. The super-resolution idea is enabled because a taxel can monitor an extended patch on the continuous surface, and multiple taxels can jointly provide information about contact information at a certain point. The particular material properties lead to a characteristic spread of contact information to the sensing taxels. Different physical effects can be used, such as electrical resistance [76], magnetic flux [15], thermo- and fluid-dynamics [143], geometric and mechanical properties [127], and so forth. The central idea is to solve the inverse problem of inferring haptic information from a few sensors, effectively creating high-resolution virtual taxels. This concept can bring high robustness, fewer wires, and less manufacturing efforts and material costs. On the other hand, even though this concept can help us design robust sensors and supply robust raw digital sensor values, these values are normally non-linear and have no direct indication of real-world values. We need real-world values for sensor implementations, such as location in millimeters and force in newtons. And this kind of haptic information is critical for the host robots to directly access the hierarchical sensing information like contact location, force strength, object curvature, hardness, and smoothness, etc., and save more resources for the robotic planning tasks. To tackle the nonlinearity problems in the data processing procedures, I develop machine-learning methods mapping raw values to real-world quantities in an end-to-end manner, and supply the host robots the needed haptic information directly, such as contact locations and force strengths.

Contribution In this thesis, I tackle three key tasks. To reduce the number of wires and create more robust sensors, I explore two strategies. One employs the super-resolution concept to acquire as much tactile information as possible from only a few physical taxels. Another applies the vision-based concept that uses a high-resolution camera with simple wiring to monitor contact surfaces from within. To process sensor data with non-linear properties, I use machine learning techniques to develop learning algorithms/frameworks for inferring real-world physical haptic information from raw sensor readings. To answer why machine learning methods can work with sensor designs, I develop a theory that links sensor measurement noises (variances) with the least-squares loss of machine learning models in a closed loop.

In the following, I compare four types of haptic sensors I designed and the data processing methods thereof from the perspectives of the sensor working mechanism (Section 7.1.1), the sensor mechanical design (Section 7.1.2), and the machine-learning data-processing procedure (Section 7.1.3).

7.1.1 Sensor Working Mechanism

I investigate several sensor design principles using different transduction methods. They share similar working mechanisms that solve the inverse problem of inferring haptic

information using machine learning (ML) from a few sensors/wire connections, effectively creating high-resolution virtual taxels. I first propose a design named **HapDef** (Chapter 2 and Chapter 3). For a large-scale stiff robotic limb, I use a few strain gauges measuring the internal deformations of the limb surface from the internal side. Since a measurement of the sensors does not directly correspond to the impacting force, an inference mechanism is required to estimate the force. I propose a data-driven approach using ML algorithms to perform this inference efficiently. Another transduction approach named **ERT-DNN** (Chapter 4) for large-scale sensing is also analyzed in this thesis. It wraps a large robotic limb surface with a whole piece of conductive fabric and assembles a few electrodes (sewed in the fabric) to read the spatial conductivity distribution in forms of voltage measurements. When an object is contacting the fabric, conductivity changes between electrodes are measured in a combinatorial manner. And the ML method interprets them into real-world haptic values, i. e., the locations and force strengths for multiple contacts.

These two approaches have both substantially reduced the amount of wires and created denser virtual taxels than actually used sensor elements/wiring connectors as needed. **HapDef** makes use of the mechanical deformation spreading behavior, and each sensor element thereof measures the local deformation information (Chapter 2 and Chapter 3). If the sensor takes local measurements, I consider the sensor has the locality property. The machine learning model trained on data with locality property has higher generalization power because a multiple-contact case would then be generalized as a combination of single contacts, similar to a superposition operation. In comparison, **ERT-DNN** depends on the electrical current flow path, and each measurement contains non-local information. The non-locality property makes the task of inverse mapping harder and increases the complexity of machine learning frameworks, according to our validation in Chapter 4. With this insight, I suggest designing the sensor considering the locality property in sensor measurements to simply the whole working mechanism.

On the other hand, rather than taking machine learning as a pure black magic box, I develop a **Theory** (Chapter 5) that gives an explanation about the relationship between the sensor measurements and machine learning models, and supplies guidance for future haptic sensor designs from different perspectives, such as transduction types, material properties, and structure designs, etc. The theory is based on sensor isolines and allows us to predict force sensitivity and accuracy in force magnitude and contact position as a spatial quantity before building the sensor. It suggests a route toward high-resolution and robust sensors that embeds a few sensor elements into a flexible surface material and uses signal processing (machine learning techniques) to achieve sensing with super-resolution accuracy. I apply the proposed theory to build a planar sensor, and the theory predicts the super-resolution very well with a coherent match to the real sensor performance. Based on this theory, I further design a 3D conical thumb-sized sensor named **BaroDome** for a manipulator end effector. I mold several barometers inside an elastomer and wrap the elastomer over a dome-shaped central core. Machine learning techniques are then used to interpret the raw barometric values into localized resultant directional force vectors.

Another sensor design concept for small-scale application is to use an internal camera

that views the soft contact surface from within. In this thesis, I also design such a vision-based haptic sensor named **Insight** (Chapter 6). It utilizes structured light and shading effects to monitor the 3D deformation of the sensing surface with a single camera. The sensor's output is computed by a data-driven machine learning approach, which directly infers distributed contact force information from raw camera readings.

BaroDome uses only a few barometers to realize super-resolution functionality, and **Insight** uses a camera centrally connected with a DAQ to realize wireless setting on the sensing surface. These two small-scale sensors both realize a few wire connections and high-resolution functionality. **BaroDome** has the potential to facilitate high-speed but less resolved studies, because there are only a few physical channels that need to be accessed. In comparison, **Insight's** core is a camera with densely integrated pixels that delivers very dense information flow trading off with speed. Depending on application requirements, different sensor design strategies can be adaptively used.

7.1.2 Sensor Mechanical Design

Most commonly used robots, especially in industry, are made of stiff/metal materials with shell structure. In those human-machine-interaction applications, haptic feedback can be easily acquired using our proposed **HapDef** design concept (Chapter 2 and Chapter 3). For robots with a thin shell, a few affordable and durable strain gauges can be directly attached to the inner side of the shell structure. We can put an additional thin shell with strain gauges assembled as the touch-sensitive skin for robots with a thick shell. Due to this design, the system is very robust to environmental impacts, and the sensor doesn't suffer from the wear problem.

ERT-DNN is another option, which wraps all the robot outer surfaces with a whole piece of durable fabric sensor Chapter 4. This design brings several conceptual advantages. Functional fabric material is very affordable nowadays. Sensors made thereof are not only soft and scalable, but also durable and easy-to-manufacture. Conductive wires can be directly sewed in the fabric material. On the one hand, fabric sensors' softness ensures the compliance to the robot surface, especially for those with curved surface. On the downside, the fabric can not recover to its original shape under strong stretching condition, which needs an extra calibration procedure over time. In addition, **ERT-DNN** also has to deal with the wear problem due to direct interaction with external environment.

Soft materials show promises in designing robotic parts (actuators, sensors, substrates) because of the favorable properties of soft materials for manipulating objects, for safer interactions in human environments, and to limit the instantaneous impact forces caused by unforeseen collisions in robotic systems. This brings both advantages and disadvantages. Due to the low Young's modulus, a sensor made thereof can be designed with very high sensitivity. However, they cannot withstand larger interaction forces, even gravity and inertial effects change their shape considerably. In addition, the nonlinear and hysteresis behaviors are more obvious than in stiffer materials, which need extra processing procedures.

Our **BaroDome** (Chapter 5) and **Insight** (Chapter 6) make a trade-off between pure soft and pure stiff materials by designing hybrid structures. These two sensors are designed for manipulation purposes, where a soft, compliant contact surface is needed. **BaroDome** is made of soft elastomer with high elongation ratio and is attached to a rigid core to support the soft material. A few barometers are sparely distributed inside the elastomer to measure spread deformation. This design ensures sensitivity (elastomer), durability (a few wires), and low cost in manufacturing procedure. However, the commercial barometers are made of metal materials, and thus have strong stiffness in contrast with the used elastomer. If the sensor is under strong interaction forces, the elastomer will be damaged from the internal side due to stress concentration effect caused by the sharp edge structure of the barometers, which needs special care.

Insight has a soft-stiff hybrid structure using over-molding. The structure is composed of two parts. One is a flexible elastomer to sense the contact, and the other is a skeleton made of aluminum alloy to support the sensing surface. In this way, the sensor is not only structurally stable, so it keeps its overall shape under high contact forces, but also sensitive, so that gentle interaction forces cause local deformations. Moreover, the sensor's shell is hollow so that the entire system is lightweight; avoiding direct contact with any optical elements also reduces the chances of image distortion and system damage. Constructing a single elastomer layer that serves all purposed is a simple, compact, robust, and wireless solution for haptic sensing. On the downside, the over-molding brings inhomogeneity in the sensor design and has a slight influence on the sensor performance, which can be tackled by machine learning procedures to some degree.

7.1.3 Machine Learning

Machine learning methods show high adaptability in solving nonlinear problems and extracting useful information from various data types. On the one hand, as discussed before, different sensors powered by individual transduction methods have different data properties. Machine learning frameworks can be specialized to cope with them. On the other hand, machine earning models are trained using data collected by automatized test beds. We have the flexibility to choose which type of data to collect and which type of data we set as the targets. Two key aspects can be categorized regarding the machine learning procedure: machine learning model and data. From the model-centric view, we collect what data we can, and develop a model good enough to deal with the noise in the data. We hold the data fixed and iteratively improve the code/model. From the data-centric view, the consistency of the data is paramount. We can use tools to improve the data quality; this allows multiple models to work well. We hold the code fixed and iteratively improve the data. In this thesis, I jointly consider these two aspects: data collection and machine learning model.

Data collection To make machine-learning-driven haptic sensors work properly, we need to collect data that ideally capture the whole spectrum of possible contact patterns. The data can be acquired by an automatized testbed or synthesized using finite element methods (FEM). The only requirement is that there should be a unique sensor reading from the sensor for a certain contact pattern. For real applications, we can purely use real collected data by the testbed to train a machine learning model to extract haptic information. It is beneficial as the collected data is automatically calibrated, while less ideal as we might not cover all contact cases, such as multiple contacts. Because performing data collection for multiple contacts is physically challenging. Multiple independent force-tips operated by more than one robot arm would be needed to stimulate the sensor in a controlled way. In comparison, we can collect any data using FEM to simulate the sensor reaction. However, the simulation model needs careful calibration and takes longer than the real data collection because accurate finite element models are computationally slow. To make a trade-off, we can establish a real2sim/sim2real framework to jointly infer contact information using data from both simulation and real physical environments (Chapter 3 and Chapter 4). The data collected in simulation environments can make up for the shortcomings of data insufficiency, which is non-trivially accessible in real simple physical environments.

Machine learning model Data from different sensors have different properties w.r.t. variants in sensor transduction methods and contact objects, such as measurement locality and sparsity. Locality describes a phenomenon that external contact events only cause value changes in taxels near the contact area. In our observation, machine learning models work less well for sensors with non-local measurement responses. For example, the machine learning model trained on single contact data can not perform well in multiple contacts prediction if we can not superpose the measurement signals of multiple single contacts together. The typical generalization problem appeared in machine learning models, which in our case is mainly caused by sensors having non-local measurement properties. We solve this problem that appeared in Chapter 4 by collecting copious good data, establishing a real2sim transfer learning framework to combine both real and simulation data, and optimizing the machine learning structure. I am now also interested in investigating whether we can bring other prior knowledge of the sensor, such as the geometric morphology, to solve the problem. I face another major problem called sparsity in training machine learning models when predicting a force map over the whole sensing surface for multiple contacts detection. Sparsity is a relative concept that compares the object contact area with the whole sensor size. For example, if the force map has a size of 100 mm $\times$ 100 mm, and the external contact only causes 1 mm $\times$ 1 mm measurement change, the machine learning model can hardly capture this sparse signal. This thesis solves it by approximating the contact force profile with an extended contact area (Chapter 4 and Chapter 6). However, other technologies can also be employed, such as active perception by sub-sampling a smaller sensor area as a target of interest.

Theory bridging data and model Moreover, the **Theory** proposed in this thesis (Chapter 5) bridges these two aspects: data and model. It introduces the concept of taxel value isolines that is an interesting way to characterize sensor responses, expressing the properties of the sensor itself and those of the environment it's placed in. The decisive quantity of the theory is the spatial distribution of the uncertainty of contact location and force magnitude inference induced by the sensor measurement noise (data noise). I aim to develop machine learning models to maximize the likelihood of predictions, which is equivalent to the search for a solution for a machine learning model with the least-squares loss. The theory describes an upper limit for the super-resolution capabilities under the model assumptions, and machine learning models can be optimized to approach this upper limit. We can use the theory to evaluate the collected data and the trained machine-learning model separately, as we expect the final sensor performance to match the theory prediction.

7.2 Outlook

With the explosion of ML technologies in broad research areas, data-driven approaches can always be invented or transferred to process data with different properties. This thesis focuses on the application of ML in haptic sensor designs. The haptic sensor design is a highly interdisciplinary research area that is still far behind its biological counterparts. A systematic design of haptic sensors involves mechanics, structure, materials, manufacturing, data acquisition, data processing, etc. Special care should be given to the fact of direct physical interaction between sensors and external objects. To ensure the long usability of haptic sensors, they need to be highly durable and robust, and the sensing characteristics should ideally not change with wear. Moreover, during sensor design phases, we need to jointly consider the sensitivity, operation range, spatial and temporal resolution, size, weight, the complexity of manufacturing, available materials in the market, structure strategies, etc. I will collect my insights from my experiences with these aspects in the following.

Material New functional materials are emerging and evolving quickly; however, their applications are comparably slow. I suggest developing haptic sensors using soft materials with low Young's modules for high sensitivity in different directions. On the downside, soft-material sensors may deform considerably under gravity or inertial effect. I use Ecoflex from Smooth-On in this thesis most frequently, as it is very durable in actual applications. Sensors made of this material have high friction on sensing surfaces, suitable for manipulation tasks. In Chapter 5, I implement a systematic analysis of material properties using finite element methods. And it shows evidence that materials with a high Poisson's ratio would be better for high sensitivity in the shear direction. In addition to mechanical properties, other properties such as optical ones used in Chapter 6 are also interesting when we consider using different transduction methods. We can use soft

materials as substrates and add additional materials to modify their optical, magnetic, or electrical properties.

Structure Design We conceive the transduction methods, select available materials, and design physical sensors accordingly to fulfill different application requirements. With a determined shape and size, I suggest considering a soft, compliant, and homogeneous sensing surface, less usage of multi-layered structure bonded by adhesives. There exists potential robustness issues that multi-layered structures bonded using adhesives can fall apart or peel off from each other under strong shear forces. A multi-layered structure is only advisable for introducing multiple functionalities and creating anisotropic properties. Some morphological designs like rigid structure are helpful to inform anisotropic properties, such as telling normal and shear contacts apart. We should jointly consider the sensor structure and manufacturing complexity for developing sensors.

Data Collection Three commonly approaches are used to collect labeled sensor data: a test bed, a fully integrated robotic system, and synthesis using finite element methods. I suggest considering the complexity of building a physical test bed, especially for those complex application scenarios with requirements of copious data. Ultimately, we would like robotic systems to do continuous learning and self-calibration. So far, we are still far away from this. Life-long exploration and exploitation inside a robotic system are recommended. These systems can acquire data labels during exploration phase and later exploit this knowledge directly when revisit them. I suggest considering the precision requirements when we refer to finite element methods. Data synthesized thereof have brought assumptions and imprecise parametrization.

Data Processing There are typically three ways to characterize sensor readings into real-world values: an analytically approximated function, a lookup table, and a machine learning model. I suggest considering the labor load and equipment complexity in data collection and calibration, the coverage of the data types, and the data storage space. Analytical solutions don't need enormous amount of data and can still have a guarantee of accuracy to some degree. However, this needs heavy hand-crafted calibration efforts. Lookup tables can help to label the sensor readings only when the data are included in the table. For wide applications, a lookup table needs very big storage space. Similarly, machine learning methods rely on copious data. However, machine learning models can be trained using the collected data, and the data do not need a big permanent storage space once the training procedure is finished. Machine learning methods are especially suited to systems with governing functions that are unknown, impractical, or impossible or represent analytically, or computationally intractable to integrate into real-world solutions [56].

In conclusion, haptic sensor design is an emerging interdisciplinary area. A one-sentence suggestion is not advisable. We need to consider different perspectives, such as

materials, structure designs, data collection, and data processing, and make suggestions for specific application scenarios. These perspectives suggest several avenues for future work. This thesis covers four robust ML-driven haptic sensors for different applications and supplies a unique theory to guide sensor designs at all levels. Optimization and application of these four types of ML-driven haptic sensors are prioritized (**HapDef** in Chapter 3, **ERT-DNN** in Chapter 4, **BaroDome** in Chapter 5, and **Insight** in Chapter 6). An application of the **Theory** is of direct interest, where the study in Chapter 5 shows a direct application in developing accelerometer-based haptic sensors. Moreover, it is interesting to broaden the coverage of my proposed **Theory** that would be beneficial to offer design guidance for different haptic sensor types.

I am asking myself: is the action we take to please our perception, or is the perception used to promote action? Where is the initial starting point, and where is the ending point?

Bibliography

[1] K. Shah, G. Ballard, A. Schmidt, and M. Schwager, "Multidrone Aerial Surveys of Penguin Colonies in Antarctica," *Science Robotics*, vol. 5, no. 47, 2020.

[2] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning Quadrupedal Locomotion over Challenging Terrain," *Science Robotics*, vol. 5, no. 47, 2020.

[3] J. Ichnowski, Y. Avigal, V. Satish, and K. Goldberg, "Deep Learning Can Accelerate Grasp-optimized Motion Planning," *Science Robotics*, vol. 5, no. 48, 2020.

[4] S. Jain, B. Thiagarajan, Z. Shi, C. Clabaugh, and M. J. Matarić, "Modeling Engagement in Long-term, In-home Socially Assistive Robot Interventions for Children with Autism Spectrum Disorders," *Science Robotics*, vol. 5, no. 39, 2020.

[5] J. C. Yang, J. Mun, S. Y. Kwon, S. Park, Z. Bao, and S. Park, "Electronic Skin: Recent Progress and Future Prospects for Skin-Attachable Devices for Health Monitoring, Robotics, and Prosthetics," *Advanced Materials*, vol. 31, no. 48, p. 1904765, 2019.

[6] J. A. Fishel and G. E. Loeb, "Sensing Tactile Microvibrations with the BioTac — Comparison with Human Sensitivity," in *Proceedings of the IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012, pp. 1122–1127.

[7] H. Lee, H. Park, G. Serhat, H. Sun, and K. J. Kuchenbecker, "Calibrating a Soft ERT-Based Tactile Sensor with a Multiphysics Model and Sim-to-real Transfer Learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 1632–1638.

[8] M. Chen, W. Luo, Z. Xu, X. Zhang, B. Xie, G. Wang, and M. Han, "An Ultrahigh Resolution Pressure Sensor Based on Percolative Metal Nanoparticle Arrays," *Nature Communications*, vol. 10, no. 1, p. 4024, Sep 2019.

[9] H. Sun and G. Martius, "Machine Learning for Haptics: Inferring Multi-Contact Stimulation From Sparse Sensor Configuration," *Frontiers in Neurorobotics*, vol. 13, p. 51, 2019.

[10] C. M. Boutry, M. Negre, M. Jorda, O. Vardoulis, A. Chortos, O. Khatib, and Z. Bao, "A Hierarchically Patterned, Bioinspired E-skin Able to Detect the Direction of Applied Pressure for Robotics," *Science Robotics*, vol. 3, no. 24, 2018.

[11] P. Mittendorfer and G. Cheng, "Humanoid Multimodal Tactile-Sensing Modules," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 401–410, 2011.

[12] J. Rogelio Guadarrama-Olvera, F. Bergner, E. Dean, and G. Cheng, "Enhancing Biped Locomotion on Unknown Terrain Using Tactile Feedback," in *Proceedings of the IEEE International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 1–9.

[13] J. Park, M. Kim, Y. Lee, H. S. Lee, and H. Ko, "Fingertip Skin–Inspired Microstructured Ferroelectric Skins Discriminate Static/Dynamic Pressure and Temperature Stimuli," *Science Advances*, vol. 1, no. 9, 2015.

[14] Y.-C. Lai, Y.-C. Hsiao, H.-M. Wu, and Z. L. Wang, "Waterproof Fabric-Based Multifunctional Triboelectric Nanogenerator for Universally Harvesting Energy from Raindrops, Wind, and Human Motions and as Self-Powered Sensors," *Advanced Science*, vol. 6, no. 5, p. 1801883, 2019.

[15] Y. Yan, Z. Hu, Z. Yang, W. Yuan, C. Song, J. Pan, and Y. Shen, "Soft Magnetic Skin for Super-resolution Tactile Sensing with Force Self-decoupling," *Science Robotics*, vol. 6, no. 51, 2021.

[16] P. Piacenza, K. Behrman, B. Schifferer, I. Kymissis, and M. Ciocarlie, "A Sensorized Multicurved Robot Finger With Data-Driven Touch Sensing via Overlapping Light Signals," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 5, pp. 2416–2427, 2020.

[17] H. Bai, S. Li, J. Barreiros, Y. Tu, C. R. Pollock, and R. F. Shepherd, "Stretchable Distributed Fiber-optic Sensors," *Science*, vol. 370, no. 6518, pp. 848–852, 2020.

[18] W. Yuan, S. Dong, and E. H. Adelson, "GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force," *Sensors*, vol. 17, no. 12, p. 2762, Nov 2017.

[19] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense Tactile Force Estimation using GelSlim and Inverse FEM," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5418–5424.

[20] C. Sferrazza and R. D'Andrea, "Design, Motivation and Evaluation of a Full-Resolution Optical Tactile Sensor," *Sensors*, vol. 19, no. 4, p. 928, Feb 2019.

[21] L. Van Duong, R. Asahina, J. Wang, and V. A. Ho, "Development of a Vision-Based Soft Tactile Muscularis," in *Proceedings of the IEEE International Conference on Soft Robotics (RoboSoft)*, 2019, pp. 343–348.

[22] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, "The TacTip Family: Soft Optical Tactile Sensors with 3D-Printed Biomimetic Morphologies," *Soft Robotics*, vol. 5, no. 2, pp. 216–227, Apr 2018.

[23] B. Lee, J.-Y. Oh, H. Cho, C. W. Joo, H. Yoon, S. Jeong, E. Oh, J. Byun, H. Kim, S. Lee, J. Seo, C. W. Park, S. Choi, N.-M. Park, S.-Y. Kang, C.-S. Hwang, S.-D. Ahn, J.-I. Lee, and Y. Hong, "Ultraflexible and Transparent Electroluminescent Skin for Real-time and Super-resolution Imaging of Pressure Distribution," *Nature Communications*, vol. 11, no. 1, p. 663, Jan 2020.

[24] M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, and W. Zaremba, "Learning Dexterous In-hand Manipulation," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 3–20, 2020.

[25] A. Nagabandi, K. Konoglie, S. Levine, and V. Kumar, "Deep Dynamics Models for Learning Dexterous Manipulation," in *Proceedings of the Conference on Robot Learning (CoRL)*, 2019.

[26] R. Bischoff, J. Kurth, G. Schreiber, R. Koeppe, A. Albu-Schaeffer, A. Beyer, O. Eiberger, S. Haddadin, A. Stemmer, G. Grunwald, and G. Hirzinger, "The KUKA-DLR Lightweight Robot Arm – a New Reference Platform for Robotics Research and Manufacturing," in *Proceedings of the International Symposium on Robotics (ISR)*, 2010, pp. 1–8.

[27] "Franks Emika: Panda," 2018.

[28] "Bosch Rexroth: APAS," 2014.

[29] G. Bledt, M. J. Powell, B. Katz, J. Di Carlo, P. M. Wensing, and S. Kim, "MIT Cheetah 3: Design and Control of a Robust, Dynamic Quadruped Robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2245–2252.

[30] "ANYbotics: ANYmal C," 2019.

[31] "Boston Dynamics: Spot and Atlas," 2019.

[32] "Shadow Robot: Deterous Hand," 2020.

[33] "Festo: TentacleGripper," 2015.

[34] M. Wehner, R. L. Truby, D. J. Fitzgerald, B. Mosadegh, G. M. Whitesides, J. A. Lewis, and R. J. Wood, "An Integrated Design and Fabrication Strategy for Entirely Soft, Autonomous Robots," *Nature*, vol. 536, no. 7617, pp. 451–455, Aug 2016.

[35] R. Deimel and O. Brock, "A Novel Type of Compliant and Underactuated Robotic Hand for Dexterous Grasping," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 161–185, 2016.

[36] F. Schmitt, O. Piccin, L. Barbé, and B. Bayle, "Soft Robots Manufacturing: A Review," *Frontiers in Robotics and AI*, vol. 5, p. 84, 2018.

[37] A. Albu-Schaffer, O. Eiberger, M. Grebenstein, S. Haddadin, C. Ott, T. Wimbock, S. Wolf, and G. Hirzinger, "Soft Robotics," *IEEE Robotics Automation Magazine*, vol. 15, no. 3, pp. 20–30, 2008.

[38] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A Review of Tactile Information: Perception and Action Through Touch," *IEEE Transactions on Robotics*, pp. 1–16, 2020.

[39] M. L. Hammock, A. Chortos, B. C.-K. Tee, J. B.-H. Tok, and Z. Bao, "25th Anniversary Article: The Evolution of Electronic Skin (E-Skin): A Brief History, Design Considerations, and Recent Progress," *Advanced Materials*, vol. 25, no. 42, pp. 5997–6038, 2013.

[40] M. Raibert, "An All Digital VLSI Tactile Array Sensor," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, 1984, pp. 314–319.

[41] A. Bicchi, J. K. Salisbury, and D. L. Brock, "Contact Sensing from Force Measurements," *The International Journal of Robotics Research*, vol. 12, no. 3, pp. 249–262, 1993.

[42] E. Cheung and V. J. Lumelsky, "Proximity Sensing in Robot Manipulator Motion Planning: System and Implementation Issues," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 6, pp. 740–751, 1989.

[43] D. Um, B. Stankovic, K. Giles, T. Hammond, and V. Lumelsky, "A Modularized Sensitive Skin for Motion Planning in Uncertain Environments," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 1, 1998, pp. 7–12 vol.1.

[44] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile Sensing—From Humans to Humanoids," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 1–20, 2010.

[45] W. Taube Navaraj, C. García Núñez, D. Shakthivel, V. Vinciguerra, F. Labeau, D. H. Gregory, and R. Dahiya, "Nanowire FET Based Neural Element for Robotic Tactile Sensing Skin," *Frontiers in Neuroscience*, vol. 11, p. 501, 2017.

[46] Y. Shao, V. Hayward, and Y. Visell, "Compression of Dynamic Tactile Information in the Human Hand," *Science Advances*, vol. 6, no. 16, 2020.

[47] L. Van Duong, R. Asahina, J. Wang, and V. A. Ho, "Development of a Vision-Based Soft Tactile Muscularis," in *Proceedings of the IEEE International Conference on Soft Robotics (RoboSoft)*, 2019, pp. 343–348.

[48] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense Tactile Force Estimation using GelSlim and Inverse FEM," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5418–5424.

[49] K. Sato, K. Kamiyama, H. Nii, N. Kawakami, and S. Tachi, "Measurement of Force Vector Field of Robotic Finger Using Vision-based Haptic Sensor," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2008, pp. 488–493.

[50] Y. Narang, K. V. Wyk, A. Mousavian, and D. Fox, "Interpreting and Predicting Tactile Signals via a Physics-Based and Data-Driven Framework," in *Proceedings of Robotics: Science and Systems (RSS)*, July 2020.

[51] A. Molchanov, O. Kroemer, Z. Su, and G. S. Sukhatme, "Contact Localization on Grasped Objects Using Tactile Sensing," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 216–222.

[52] B. Navarro, P. Kumar, A. Fonte, P. Fraisse, G. Poisson, and A. Cherubini, "Active Calibration of Tactile Sensors Mounted on a Robotic Hand," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2015.

[53] Z. Su, J. Fishel, T. Yamamoto, and G. Loeb, "Use of Tactile Feedback to Control Exploratory Movements to Characterize Object Compliance," *Frontiers in Neurorobotics*, vol. 6, p. 7, 2012.

[54] M. Lee and H. Nicholls, "Review Article Tactile Sensing for Mechatronics – a State of the Art Survey," *Mechatronics*, vol. 9, pp. 1–31, 1999.

[55] Z. Li, S. Liu, and C.-Y. Su, "A Novel Analytical Inverse Compensation Approach for Preisach Model," in *Proceedings of the International Conference on Intelligent Robotics and Applications*, 2013, pp. 656–665.

[56] K. Chin, T. Hellebrekers, and C. Majidi, "Machine Learning for Soft Robotic Sensing and Control," *Advanced Intelligent Systems*, vol. 2, no. 6, p. 1900171, 2020.

[57] Z. Su, K. Hausman, Y. Chebotar, A. Molchanov, G. E. Loeb, G. S. Sukhatme, and S. Schaal, "Force Estimation and Slip Detection/Classification for Grip Control Using a Biomimetic Tactile Sensor," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2015, pp. 297–303.

[58] T. Horii, Y. Nagai, L. Natale, F. Giovannini, G. Metta, and M. Asada, "Compensation for Tactile Hysteresis Using Gaussian Process with Sensory Markov Property," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2014, pp. 993–998.

[59] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent Hardness Estimation using Deep Learning and a GelSight Tactile Sensor," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 951–958.

[60] W. S. McCulloch and W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec 1943.

[61] G. L. Shaw, "Donald Hebb: The Organization of Behavior," in *Brain Theory*, 1986, pp. 231–233.

[62] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[63] P. J. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. dissertation, Harvard University, 1974.

[64] J. R. Quinlan, "Induction of Decision Trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar 1986.

[65] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[66] R. E. Schapire, "The Strength of Weak Learnability," *Machine Learning*, vol. 5, no. 2, pp. 197–227, Jun 1990.

[67] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep 1995.

[68] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct 2001.

[69] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[70] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[71] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[72] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile Sensing in Dexterous Robot Hands – Review," *Robotics and Autonomous Systems*, vol. 74, pp. 195–220, 2015.

[73] M. T. Francomano, D. Accoto, and E. Guglielmelli, "Artificial Sense of Slip—A Review," *IEEE Sensors Journal*, vol. 13, no. 7, pp. 2489–2498, 2013.

[74] R. S. Dahiya, P. Mittendorfer, M. Valle, G. Cheng, and V. J. Lumelsky, "Directions Toward Effective Utilization of Tactile Skin: A Review," *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4121–4138, 2013.

[75] R. Balasubramanian and V. J. Santos, *The Human Hand as an Inspiration for Robot Hand Development*. Springer Publishing Company, Incorporated, 2014.

[76] H. Lee, D. Kwon, H. Cho, I. Park, and J. Kim, "Soft Nanocomposite Based Multi-point, Multi-directional Strain Mapping Sensor Using Anisotropic Electrical Impedance Tomography," *Scientific Reports*, vol. 7, no. 1, p. 39837, Jan 2017.

[77] R. Kõiva, M. Zenker, C. Schürmann, R. Haschke, and H. J. Ritter, "A Highly Sensitive 3D-shaped Tactile Sensor," in *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2013, pp. 1084–1089.

[78] C. A. Jara, J. Pomares, F. A. Candelas, and F. Torres, "Control Framework for Dexterous Manipulation Using Dynamic Visual Servoing and Tactile Sensors' Feedback," *Sensors*, vol. 14, no. 1, pp. 1787–1804, 2014.

[79] B. Heyneman and M. R. Cutkosky, "Biologically Inspired Tactile Classification of Object-hand and Object-world Interactions," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2012, pp. 167–173.

[80] D. Goger, N. Gorges, and H. Worn, "Tactile Sensing for an Anthropomorphic Robotic Hand: Hardware and Signal Processing," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2009, pp. 895–901.

[81] L. Seminara, M. Capurro, P. Cirillo, G. Cannata, and M. Valle, "Electromechanical Characterization of Piezoelectric PVDF Polymer Films for Tactile Sensors in Robotics Applications," *Sensors and Actuators A: Physical*, vol. 169, no. 1, pp. 49–58, 2011.

[82] K. J. Kuchenbecker, J. Gewirtz, W. McMahan, D. Standish, P. Martin, J. Bohren, P. J. Mendoza, and D. I. Lee, "VerroTouch: High-Frequency Acceleration Feedback for Telerobotic Surgery," in *Haptics: Generating and Perceiving Tangible Sensations*, 2010, pp. 189–196.

[83] R. S. Johansson and J. R. Flanagan, "Coding and Use of Tactile Signals from the Fingertips in Object Manipulation Tasks," *Nature Reviews Neuroscience*, vol. 10, no. 5, pp. 345–359, May 2009.

[84] Q. Li, C. SchÃrmann, R. Haschke, and H. Ritter, "A Control Framework for Tactile Servoing," in *Proceedings of Robotics: Science and Systems (RSS)*, June 2013.

[85] E. C. Swanson, E. J. Weathersby, J. C. Cagle, and J. E. Sanders, "Evaluation of Force Sensing Resistors for the Measurement of Interface Pressures in Lower Limb Prosthetics," *Journal of Biomechanical Engineering*, vol. 141, no. 10, Jul 2019.

[86] A. S. Sadun, J. Jalani, and J. A. Sukor, "Force Sensing Resistor (FSR): a Brief Overview and the Low-cost Sensor for Active Compliance Control," in *Proceedings of the International Workshop on Pattern Recognition*, vol. 10011, 2016, pp. 222 – 226.

[87] S. Teshigawara, S. Shimizu, K. Tadakuma, M. Aiguo, M. Shimojo, and M. Ishikawa, "High Sensitivity Slip Sensor Using Pressure Conductive Rubber," *Sensors, 2009 IEEE*, pp. 988–991, 2009.

[88] H. Lee, K. Park, J. Kim, and K. J. Kuchenbecker, "Internal Array Electrodes Improve the Spatial Resolution of Soft Tactile Sensors Based on Electrical Resistance Tomography," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5411–5417.

[89] M. Inaba, Y. Hoshino, K. Nagasaka, T. Ninomiya, S. Kagami, and H. Inoue, "A Full-body Tactile Sensor Suit Using Electrically Conductive Fabric and Strings," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2, 1996, pp. 450–457.

[90] D. J. Beebe, A. S. Hsieh, D. D. Denton, and R. G. Radwin, "A Silicon Force Sensor for Robotics and Medicine," *Sensors and Actuators A: Physical*, vol. 50, no. 1, pp. 55–65, 1995.

[91] M. Wolffenbuttel and P. Regtien, "Polysilicon Bridges for the Realization of Tactile Sensors," *Sensors and Actuators A: Physical*, vol. 26, no. 1, pp. 257–264, 1991.

[92] A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta, "Methods and Technologies for the Implementation of Large-Scale Robot Tactile Sensors," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 389–400, 2011.

[93] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker, "Human-Inspired Robotic Grasp Control With Tactile Sensing," *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1067–1079, 2011.

[94] "PPS: DigiTacts," 2020.

[95] "Analog Devices: AD7147," 2020.

[96] J. R. Flanagan and A. M. Wing, "Modulation of Grip Force with Load Force During Point-to-Point Arm Movements," *Experimental Brain Research*, vol. 95, no. 1, pp. 131–143, Jul 1993.

[97] T. Zhang, H. Liu, L. Jiang, S. Fan, and J. Yang, "Development of a Flexible 3-D Tactile Sensor System for Anthropomorphic Artificial Hand," *IEEE Sensors Journal*, vol. 13, no. 2, pp. 510–518, 2013.

[98] S. Company, "Developments in Dexterous Hands for Advanced Robotic Applications," in *Proceedings of the World Automation Congress, 2004.*, vol. 15, 2004, pp. 123–128.

[99] T. B. Martin, R. O. Ambrose, M. A. Diftler, R. Platt, and M. J. Butzer, "Tactile Gloves for Autonomous Grasping with the NASA/DARPA Robonaut," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, 2004, pp. 1713–1718 Vol.2.

[100] "Peratech: Quantum Tunneling Composite," 2014.

[101] J. Fraden, *Handbook of Modern Sensors: Physics, Designs, and Applications*. SpringerVerlag, 2003.

[102] K. O. Hill, Y. Fujii, D. C. Johnson, and B. S. Kawasaki, "Photosensitivity in Optical Fiber Waveguides: Application to Reflection Filter Fabrication," *Applied Physics Letters*, vol. 32, no. 10, pp. 647–649, 1978.

[103] M. I. Tiwana, S. J. Redmond, and N. H. Lovell, "A Review of Tactile Sensing Technologies with Applications in Biomedical Engineering," *Sensors and Actuators A: Physical*, vol. 179, pp. 17–31, 2012.

[104] Y. Tenzer, L. P. Jentoft, and R. D. Howe, "The Feel of MEMS Barometers: Inexpensive and Easily Customized Tactile Array Sensors," *IEEE Robotics Automation Magazine*, vol. 21, no. 3, pp. 89–95, 2014.

[105] "TakkStrip," 2021.

[106] L. U. Odhner, L. P. Jentoft, M. R. Claffee, N. Corson, Y. Tenzer, R. R. Ma, M. Buehler, R. Kohout, R. D. Howe, and A. M. Dollar, "A Compliant, Underactuated Hand for Robust Manipulation," *The International Journal of Robotics Research*, vol. 33, no. 5, pp. 736–752, 2014.

[107] D. F. Gomes, Z. Lin, and S. Luo, "GelTip: A Finger-shaped Optical Tactile Sensor for Robotic Manipulation," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 9903–9909.

[108] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "OmniTact: A Multi-Directional High-Resolution Touch Sensor," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 618–624.

[109] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense Tactile Force Estimation using GelSlim and Inverse FEM," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5418–5424.

[110] X. Lin and M. Wiertlewski, "Sensing the Frictional State of a Robotic Skin via Subtractive Color Mixing," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2386–2392, 2019.

[111] L. Van Duong and V. A. Ho, "Large-Scale Vision-Based Tactile Sensing for Robot Links: Design, Modeling, and Evaluation," *IEEE Transactions on Robotics*, pp. 1–14, 2020.

[112] L. Van Duong, R. Asahina, J. Wang, and V. A. Ho, "Development of a Vision-Based Soft Tactile Muscularis," in *Proceedings of the IEEE International Conference on Soft Robotics (RoboSoft)*, 2019, pp. 343–348.

[113] K. Koyama, M. Shimojo, T. Senoo, and M. Ishikawa, "High-Speed High-Precision Proximity Sensor for Detection of Tilt, Distance, and Contact," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3224–3231, 2018.

[114] A. J. Smola and B. Schölkopf, "A Tutorial on Support Vector Regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, Aug. 2004.

[115] N. Cressie, *Statistics for Spatial Data*, ser. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley, 1993.

[116] C. E. Rasmussen, "Gaussian Processes for Machine Learning," in *Gaussian processes for machine learning*. MIT Press, 2006.

[117] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, 2002.

[118] W. Yuan, Y. Mo, S. Wang, and E. H. Adelson, "Active Clothing Material Perception Using Tactile Sensing and Deep Learning," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 4842–4849.

[119] Q. Yuan and J. Wang, "Design and Experiment of the NAO Humanoid Robot's Plantar Tactile Sensor for Surface Classification," in *Proceedings of the International Conference on Information Science and Control Engineering (ICISCE)*, 2017, pp. 931–935.

[120] V. Prado da Fonseca, T. E. Alves de Oliveira, and E. M. Petriu, "Estimating the Orientation of Objects from Tactile Sensing Data Using Machine Learning Methods and Visual Frames of Reference," *Sensors*, vol. 19, no. 10, 2019.

[121] W. Becari, L. Ruiz, B. G. P. Evaristo, and F. J. Ramirez-Fernandez, "Comparative Analysis of Classification Algorithms on Tactile Sensors," in *Proceedings of the IEEE International Symposium on Consumer Electronics (ISCE)*, 2016, pp. 1–2.

[122] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar, "Single-Grasp Object Classification and Feature Extraction with Simple Robot Hands and Tactile Sensors," *IEEE Transactions on Haptics*, vol. 9, no. 2, pp. 207–220, 2016.

[123] M. Lapeyre, P. Rouanet, J. Grizou, S. Nguyen, F. Depraetre, A. Le Falher, and P.-Y. Oudeyer, "Poppy Project: Open-Source Fabrication of 3D Printed Humanoid Robot for Science, Education and Art," in *Digital Intelligence 2014*, Sep. 2014, p. 6.

[124] H. Lee, K. Park, J. Kim, and K. J. Kuchenbecker, "Internal Array Electrodes Improve the Spatial Resolution of Soft Tactile Sensors Based on Electrical Resistance Tomography," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5411–5417.

[125] G. Westheimer, "Optical Superresolution and Visual Hyperacuity," *Progress in Retinal and Eye Research*, vol. 31, no. 5, pp. 467–480, 2012.

[126] A. Krause, A. Singh, and C. Guestrin, "Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies," *Journal of Machine Learning Research (JMLR)*, vol. 9, pp. 235–284, February 2008.

[127] N. F. Lepora, U. Martinez-Hernandez, M. Evans, L. Natale, G. Metta, and T. J. Prescott, "Tactile Superresolution and Biomimetic Hyperacuity," *IEEE Transactions on Robotics*, vol. 31, no. 3, pp. 605–618, 2015.

[128] Lichuan Liu, S. M. Kuo, and M. Zhou, "Virtual Sensing Techniques and Their Applications," in *Proceedings of the IEEE International Conference on Networking, Sensing and Control*, 2009, pp. 31–36.

[129] H. Sun and G. Martius, "Robust Affordable 3D Haptic Sensation via Learning Deformation Patterns," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 846–853.

[130] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[131] H. Lee, H. Sun, H. Park, G. Serhat, B. Javot, G. Martius, and K. J. Kuchenbecker, "Predicting the force map of an ert-based tactile sensor using simulation and deep networks," *IEEE Transactions on Automation Science and Engineering*, pp. 1–15, 2022.

[132] H. Sun and G. Martius, "Guiding the design of superresolution tactile skins with taxel value isolines theory," *Science Robotics*, vol. 7, no. 63, p. eabm0608, 2022.

[133] P. A. Tres, "10 - In-Mold Assembly," in *Designing Plastic Parts for Assembly*. Hanser, 2014, pp. 327 – 341.

[134] H. Sun, K. J. Kuchenbecker, and G. Martius, "A Soft Thumb-sized Vision-based Sensor with Accurate All-round Force Perception," vol. 4, no. 2, pp. 135–145, Feb 2022.

[135] M. Lapeyre, P. Rouanet, J. Grizou, S. Nguyen, F. Depraetre, A. Le Falher, and P.-Y. Oudeyer, "Poppy Project: Open-Source Fabrication of 3D Printed Humanoid Robot for Science, Education and Art," in *Digital Intelligence 2014*, Sep. 2014, p. 6.

[136] K. Lawrence, *ANSYS Tutorial Release 13*. SDC Publications, 2011.

[137] E. Candès and M. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[138] K. Manohar, B. W. Brunton, J. N. Kutz, and S. L. Brunton, "Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns," *IEEE Control Systems Magazine*, vol. 38, no. 3, pp. 63–86, 2018.

[139] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[140] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[141] I. Jolliffe, *Principal Component Analysis*. Springer Verlag, 1986.

[142] E. Ventsel and T. Krauthammer, *Thin Plates and Shells: Theory: Analysis, and Applications*. Taylor & Francis, 2001.

[143] A. Krause, A. Singh, and C. Guestrin, "Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies," *Journal of Machine Learning Research (JMLR)*, vol. 9, pp. 235–284, February 2008.

[144] D. M. Stefanescu, "Strain Gauges and Wheatstone Bridges — Basic Instrumentation and New Applications for Electrical Measurement of Non-electrical Quantities," in *Proceedings of the International Multi-Conference on Systems, Signals and Devices*, 2011, pp. 1–5.

[145] L. S. Bickley, P. G. Szilagyi, and R. M. Hoffman., *Bates' Guide to Physical Examination and History Taking*. Wolters Kluwer, 2017.

[146] D. Shooter, "Use of Two-point Discrimination as a Nerve Repair Assessment Tool: Preliminary Report," *ANZ journal of surgery*, vol. 75, pp. 866–8, 10 2005.

[147] H. Blumenfeld, *Neuroanatomy through Clinical Cases*. Sinauer Associates, 2010.

[148] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, 2006.

[149] F. A. Anifowose, "A Comparative Study of Gaussian Mixture Model and Radial Basis Function for Voice Recognition," *CoRR*, vol. abs/1211.2556, 2012.

[150] R. A. Russell and S. Parkinson, "Sensing Surface Shape by Touch," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1993, pp. 423–428.

[151] D. S. Holder, *Electrical Impedance Tomography: Methods, History and Applications*. CRC Press, 2004.

[152] D. S. Tawil, D. Rye, and M. Velonaki, "Improved image reconstruction for an EIT-based sensitive skin with multiple internal electrodes," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 425–435, 2011.

[153] D. Silvera-Tawil, D. Rye, M. Soleimani, and M. Velonaki, "Electrical Impedance Tomography for Artificial Sensitive Robotic Skin: a Review," *IEEE Sensors Journal*, vol. 15, no. 4, pp. 2001–2016, 2015.

[154] S. Martin and C. T. Choi, "Nonlinear Electrical Impedance Tomography Reconstruction Using Artificial Neural Networks and Particle Swarm Optimization," *IEEE Transactions on Magnetics*, vol. 52, no. 3, pp. 1–4, 2015.

[155] S. Martin and C. T. M. Choi, "A Post-Processing Method for Three-Dimensional Electrical Impedance Tomography," *Scientific Reports*, vol. 7, no. 1, p. 7212, Aug 2017.

[156] H. Park, K. Park, S. Mo, and J. Kim, "Deep Neural Network Based Electrical Impedance Tomographic Sensing Methodology for Large-Area Robotic Tactile Sensing," *IEEE Transactions on Robotics*, 2021.

[157] Y. Zhang, G. Laput, and C. Harrison, "Electrick: Low-cost Touch Sensing Using Electric Field Tomography," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1–14.

[158] S. Russo, R. Assaf, N. Carbonaro, and A. Tognetti, "Touch Position Detection in Electrical Tomography Tactile Sensors Through Quadratic Classifier," *IEEE Sensors Journal*, vol. 19, no. 2, pp. 474–483, 2018.

[159] D. S. Tawil, D. Rye, and M. Velonaki, "Interpretation of the Modality of Touch on an Artificial Arm Covered with an EIT-based Sensitive Skin," *The International Journal of Robotics Research*, vol. 31, no. 13, pp. 1627–1641, 2012.

[160] J. A. Rinde, "Poisson's Ratio for Rigid Plastic Foams," *Journal of Applied Polymer Science*, vol. 14, pp. 1913–1926, 1970.

[161] R. W. Williams, "Measuring and Modeling the Anisotropic, Nonlinear and Hysteretic Behavior of Woven Fabrics," Ph.D. dissertation, University of Iowa, 2010.

[162] C. Bartolozzi, L. Natale, F. Nori, and G. Metta, "Robots with a sense of touch," *Nature Materials*, vol. 15, no. 9, pp. 921–925, Sep 2016.

[163] V. Marx, "Is Super-resolution Microscopy Right for You?" *Nature Methods*, vol. 10, no. 12, pp. 1157–1163, Dec 2013.

[164] E. B. Fox, J. W. Fisher, and A. S. Willsky, "Detection and Localization of Material Releases With Sparse Sensor Configurations," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1886–1898, 2007.

[165] K. S. Eu and K. M. Yap, "Chemical Plume Tracing: A Three-dimensional Technique for Quadrotors by Considering the Altitude Control of the Robot in the Casting Stage," *International Journal of Advanced Robotic Systems*, vol. 15, no. 1, p. 1729881418755877, 2018.

[166] R. Mahmood, S. Jia, and W. Zhu, "Analysis of Climate Variability, Trends, and Prediction in the Most Active Parts of the Lake Chad Basin, Africa," *Scientific Reports*, vol. 9, no. 1, p. 6317, Apr 2019.

[167] N. F. Lepora and B. Ward-Cherrier, "Superresolution with an Optical Tactile Sensor," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 2686–2691.

[168] P. Piacenza, S. Sherman, and M. Ciocarlie, "Data-Driven Super-Resolution on a Tactile Dome," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1434–1441, 2018.

[169] T. Hellebrekers, N. Chang, K. Chin, M. J. Ford, O. Kroemer, and C. Majidi, "Soft Magnetic Tactile Skin for Continuous Force and Location Estimation Using Neural Networks," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3892–3898, 2020.

[170] T. Q. Pham, T. Hoshi, Y. Tanaka, and A. Sano, "Effect of 3D Microstructure of Dermal Papillae on SED Concentration at a Mechanoreceptor Location," *PLOS ONE*, vol. 12, no. 12, pp. 1–16, 12 2017.

[171] P. H. Mott and C. M. Roland, "Limits to Poisson's Ratio in Isotropic Materials – General Result for Arbitrary Deformation," *Physica Scripta*, vol. 87, no. 5, p. 055404, 2013.

[172] P. Dallaire, P. Giguère, D. Émond, and B. Chaib-draa, "Autonomous Tactile Perception: A Combined Improved Sensing and Bayesian Nonparametric Approach," *Robotics and Autonomous Systems*, vol. 62, no. 4, pp. 422–435, 2014.

[173] H. Lee, J. Chung, S. Chang, and E. Yoon, "Normal and Shear Force Measurement Using a Flexible Polymer Tactile Sensor With Embedded Multiple Capacitors," *Journal of Microelectromechanical Systems*, vol. 17, no. 4, pp. 934–942, 2008.

[174] M. Lambeta, P. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, and R. Calandra, "DIGIT: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor With Application to In-Hand Manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.

[175] C. Sferrazza, T. Bi, and R. D'Andrea, "Learning the Sense of Touch in Simulation: a Sim-to-real Strategy for Vision-based Tactile Sensing," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 4389–4396.

[176] R. J. Woodham, "Photometric Method For Determining Surface Orientation From Multiple Images," *Optical Engineering*, vol. 19, no. 1, pp. 139–144, 1980.

[177] J. Geng, "Structured-light 3D Surface Imaging: a Tutorial," *Advances in Optics and Photonics*, vol. 3, no. 2, pp. 128–160, Jun 2011.

[178] N. Perez, *Theory of Elasticity*. Springer International Publishing, 2017, pp. 1–52.

[179] D. Rus and M. T. Tolley, "Design, Fabrication and Control of Soft Robots," *Nature*, vol. 521, no. 7553, pp. 467–475, May 2015.

[180] A. Shahnewaz and A. K. Pandey, *Color and Depth Sensing Sensor Technologies for Robotics and Machine Vision*. Springer International Publishing, 2020, pp. 59–86.

[181] D. G. Aliaga and Yi Xu, "Photogeometric Structured Light: A Self-calibrating and Multi-viewpoint Framework for Accurate 3D Modeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[182] M. Wüthrich, F. Widmaier, F. Grimminger, J. Akpo, S. Joshi, V. Agrawal, B. Hammoud, M. Khadiv, M. Bogdanovic, V. Berenz, J. Viereck, M. Naveau, L. Righetti, B. Schölkopf, and S. Bauer, "TriFinger: An Open-Source Robot for Learning Dexterity," 2020.

[183] H. W. Kuhn and B. Yaw, "The Hungarian Method for the Assignment Problem," *Naval Research Logistics Quarterly*, pp. 83–97, 1955.

[184] K. L. Johnson, *Normal Contact of Elastic Solids – Hertz Theory*. Cambridge University Press, 1985, pp. 84—-106.

[185] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "GelSlim: A High-Resolution, Compact, Robust, and Calibrated Tactile-sensing Finger," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1927–1934.

[186] S. Dong, W. Yuan, and E. H. Adelson, "Improved GelSight Tactile Sensor for Measuring Geometry and Slip," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 137–144.

[187] A. Wilson, S. Wang, B. Romero, and E. Adelson, "Design of a Fully Actuated Robotic Hand With Multiple Gelsight Tactile Sensors," 2020.