QUEEN MARY UNIVERSITY OF LONDON

Faculty of Science and Engineering

School of Engineering and Materials Science

Photoelastic Tactile Sensor for a Robotic Palm

Staines Rajith

Student ID: 250790023

Supervised by

Prof Kaspar Althoefer

Dr Abu Bakar Dawood

Dr Dalia Osman

2025-2026

Submitted in fulfilment of the requirements for the degree of MSc at Queen Mary University of London

DECLARATION OF OWN WORK

School of Engineering and Materials Science

MSc in Robotics and AI
EMS715P Extended Research Project
AUGUST 2026

DECLARATION

This report entitled

Photoelastic Tactile Sensor for a Robotic Palm

Was composed by me and is based on my own work. Where the work of others has been used, it is fully acknowledged in the text and in captions to table illustrations. This report has not been submitted for any other qualification.
Name: Staines Rajith

Signed: *Staines Rajith*

Date: *30/08/26*

Impact Statement

This statement is provided to help examiners understand the impact that the laboratory fire incident and subsequent access restrictions may have had on the conduct of this project, as well as any measures taken to mitigate those effects.

---

1. Impact of the incident on the project

Briefly describe how the fire incident affected your project (e.g., loss of laboratory access, equipment unavailability, restricted access to facilities, loss of materials, interruption to experiments).

Response:

A fire occurred in the laboratory on the day of the main experimental data collection. Normal indentation was completed, but access to the lab and experimental equipment was limited afterwards. This halted further experimental work on the rig and facilities for the remainder of the project. Thus, the data obtained just before the disruption were used as the primary experimental dataset for sensor calibration and assessment.

---

2. Impact on planned research activities

Please explain which planned research activities could not be completed, were delayed, or were only partially completed as a result of the disruption.

Response:

The original experimental programme was planned to involve intentional lateral loading, as well as experiments with different indenter geometries beyond the normal indentation. The following experiments were designed to get specific calibration samples for 3-axis force measurement and contact-shape classification.

Controlled normal indentation was mainly used in the finished experiment. The ATI force/torque sensor measured F_x, F_y, and F_z during each indentation, but in the experiment, the lateral force was neither applied nor controlled. The recorded F_x and F_y components are thus a set of forces generated by the normal-indentation procedure,

not a separate set of shear forces. The planned experiments with controlled lateral loading and differently shaped indenters could not be repeated due to the loss of access to the laboratory and the required experimental facilities following the fire.
As the fire occurred at an early stage of the experimental work, I was also unable to collect additional photographs and diagrams of the experimental setup and procedures that would have supported the descriptions presented in the report. The disruption also meant that an additional independent dataset was not collected after model development. Because of this, the evaluation had to be conducted using partitions of the dataset created during the completed indentation experiment, rather than in a new experimental campaign with new contacts or objects.

---

3. Mitigation measures adopted

Please describe any actions taken to mitigate the impact of the disruption (e.g., revised experimental plan, use of alternative facilities, additional literature review, modelling, simulation, secondary data analysis).

Response:
After the disruption, the project was re-scoped around the experimental data which had been successful to date. To avoid replicating the experimental setup without access to the necessary facilities, the remaining project time was used to explore several ways of calibrating using the available data and various image representations. Four machine learning architectures were tested, including raw and differential image representations, convolutional networks, a flattened-intensity MLP, and a two-stage model in which the depth of indentation and contact location were estimated first, followed by force estimation.
The data were recorded, and repeated measurements were conducted across the surface to be sensed; multiple indentation depths were tested; synchronised camera frames were captured; known contact coordinates were used; and the force was measured with the ATI sensor. This enabled the completed normal-indentation experiment to be used to compare more calibration strategies than originally intended.
Physical location leakage of training and evaluation was avoided. I used three convolutional architectures by applying `GroupShuffleSplit` to group samples by `point_id` and ensure that each physical contact location was assigned to a single partition.

Samples were grouped by physical grid coordinates (row, col), and the flattened-array MLP was derived by calling the function `build_group_split`. Evaluation samples were then sampled from contact locations not used for training, testing predictions at data locations not used for training.

---

4. Changes to methodology or project scope

Please outline any changes made to the original methodology, research design, objectives, or project scope following the disruption.

Response:

The project scope was changed from a full calibration of normal and lateral components of force and a classification of the shape of the contact to one based primarily on estimation of the normal force, contact localisation and estimation of indentation depth, to analyse the data acquired for the normal component of force, which had already been successfully obtained.

All three of the force components were measured by the ATI force/torque sensor, but F_z was only intentionally changed during the controlled indentation process. Thus, the result of the force calibration is assumed to be the principal force calibration result F_z. Recorded values of F_x and F_y are used for predictions and stored as further observations of the models' behaviour; however, they are not considered a validation of controlled shear-force sensing.

Similarly, contact-shape classification was excluded from the experimental dataset because only one indenter shape was available. The available data were instead used to explore alternative machine-learning architectures and the impact of using the raw, differential, RGB, greyscale, convolutional and flattened-intensity image representations.

---

5. Resulting limitations

Please identify any limitations arising from the incident and explain how these may have affected the project's findings, conclusions, or recommendations.

Response:

The principal drawback due to disruption is the reduced scope of the experimental validation. Calibration is shown primarily for controlled normal indentation using a single-indenter geometry. It does not provide experimental proof of general three-axis force sensing and contact-shape classification because dedicated datasets of lateral loading and multi-shape were not collected.

Another caveat is that generalisations beyond the completed experimental campaign are limited. All four architectures used location-disjoint partitions; that is, physical locations in the evaluation set were not used for training. However, the training/evaluation sets were still from the same sensor, experimental setup, and data-collection campaign. The test results are then generalised to a held-out set of contact locations in the reported experimental distribution, but do not necessarily provide evidence of generalisation to unseen objects, independently collected datasets, different sensor builds, or significantly different operating conditions. This is especially pertinent to the contact-localisation results.

Hence, the reported results are to be seen as calibration and proof-of-concept testing of the proposed sensing architecture, and not as a full test of the architecture under arbitrary contact conditions. Additional experiments with independently gathered data, new objects, controlled lateral loading, different indenter geometries, and realistic manipulations are still to be carried out to grant further generalisation and complete multi-axis sensing.

Student Declaration

I confirm that this statement provides an accurate account of the impact of the laboratory fire incident on my MSc project and the actions taken in response.

Signature: *Staines Rajith* Date: *30/08/2026*

Declaration of Generative AI Use

Generative artificial intelligence (AI) tools were used during the preparation of this dissertation solely to assist with grammar checking and the rephrasing of text originally written by the author. AI tools were not used to generate original sentences, arguments, research findings, experimental results, or scientific conclusions. Their use in the written dissertation was limited to improving the grammar, clarity, and phrasing of the author's existing text.

Generative AI tools were also used to assist with aspects of code development, including code modification, and debugging. The resulting code was reviewed, tested, and adapted by the author as required for the methods and analyses presented in this dissertation. The source code used for the project is available in the public GitHub repository referenced in Appendix A.1.

I confirm that all AI-assisted material was reviewed by me and that I take full responsibility for the accuracy, interpretation, and final content of this dissertation.

Signature: *Staines Rajith* Date: *30/08/2026*

Abstract

Tactile sensing is essential for robotic manipulation to detect contact forces and locations during physical interaction. The incorporation of tactile sensors into robotic hands is challenging, as the sensors must be small enough to fit within the hand without limiting its motion or function. The palm is particularly relevant for tactile sensing because it provides a large contact area that can help support and stabilise objects during manipulation. But when tactile sensing is distributed across the entire palm, there are also constraints on sensor size and optical configuration. While vision-based tactile sensing can provide rich spatial information, many existing approaches rely on optical configurations that are difficult to integrate within a compact robotic palm.

This work introduces a low-cost, vision-based tactile sensor for robotic palms that uses the photoelastic effect to measure contact force and position by analysing fringe patterns. The optical path is bent beneath the sensing membrane by an angled mirror, allowing the camera to be positioned near the wrist and reducing the space beneath the palm. A computer-controlled Cartesian stage was programmed to simultaneously collect a set of fringe images, contact position, indentation depth, and three-axis force. Four architectures for raw and differential image processing were tested: a joint multi-task network, a force-conditioned network, a multi-layer perceptron applied directly to flattened pixel arrays, and a physics-informed two-stage network comprising an indentation depth- and location-estimation network followed by a force-regression network. The physics-informed two-stage network achieved sub-millimetre contact localisation with coordinate MAEs of 0.131 mm and 0.147 mm, and $R^2 > 0.999$. The lowest normal force error of 0.244 N was obtained for the differential RGB model. The same pipeline achieved an indentation-depth MAE of 0.165 mm and $R^2 = 0.978$; the lightweight multilayer perceptron was competitive, with a force prediction MAE of 0.330 N. The results presented here highlight the potential of photoelastic fringe patterns to provide a high degree of optical information for tactile sensing at the palm scale, and show that a compact, modular optical system with relatively lightweight calibration models can achieve accurate force and contact localisation under the conditions evaluated here.

Keywords: Photoelasticity, tactile sensing, optical force estimation, contact localisation, palm-scale sensing, vision-based tactile sensing, machine learning

Contents

Impact Statement **1**

Declaration of Generative AI Use **5**

Abstract **6**

List of Figures **9**

List of Tables **10**

Nomenclature **11**

1 Introduction **13**

2 Background and State-of-the-Art **15**

2.1 Vision-Based Tactile Sensing for Robotic Manipulation 15

2.2 Large-Area and Palm-Scale Vision-Based Tactile Sensors 16

2.3 Photoelasticity-Based Tactile Sensing 17

2.4 Deep Learning and Calibration Pipelines 18

2.5 Research Gap and Motivation for a Photoelastic Tactile Palm 19

3 Principles of Photoelastic Tactile Sensing **22**

4 Sensor Design and Fabrication **24**

4.1 Material Selection 24

4.2 Prototyping 25

4.3 Diffused Back-Illumination and Folded Optical Path 27

5 Experimental Setup and Calibration **28**

5.1 Experimental Setup 28

5.2 Model Architecture and Training 32

5.2.1 Joint Multi-Task Network with Raw Crop Inputs 34

5.2.2 Force-Conditioned Network with Differential Crop Inputs 35

5.2.3 MLP Regression on Flattened 1D Delta-Crop Arrays 36

5.2.4 Physics-Informed Two-Stage Estimation Pipeline 36

6 Results **38**

6.1 Evaluation Metrics 38

6.2 Quantitative Results 39

6.2.1 Force Prediction . . . 39
6.2.2 Contact-Location Prediction . . . 41
6.2.3 Comparison of Calibration Architectures . . . 43

7 Comparison with State-of-the-Art . . . 45

8 Discussion . . . 47
8.1 Key Findings and Contributions . . . 47
8.2 Fringe Quality and Optical Path Design . . . 47
8.3 Design and Deployment Trade-offs . . . 47
8.4 Calibration Architecture and Representation Choice . . . 48
8.5 Advantages . . . 48
8.6 Limitations . . . 49

9 Conclusion . . . 50

10 Future Work . . . 51

References . . . 55

Acknowledgements . . . 56

A Appendix . . . 57
A.1 Code Repository . . . 57

List of Figures

1 Initial prototype 25
2 Final prototype: (a) exploded view showing the discrete illumination, polarisation, and mounting components; (b) assembled view. 26
3 Exploded view of the back-illumination assembly 27
4 Sensor mounted on the computer-controlled Cartesian stage, with the ATI Mini force/torque sensor and indenter positioned above. The camera is secured to the white mount, viewing the sensor cross-section, with a lid enclosing the optical path to minimise ambient-light interference. 28
5 Camera view of the sensing membrane under no-contact and applied-contact conditions. 29
6 Annotated camera view under applied contact. 29
7 Camera views illustrating changes in the photoelastic response with indentation depth and contact location. Panels (a) and (b) show the same contact location at 1 mm and 5 mm indentation depths, respectively, while panels (b) and (c) show different contact locations at the same 5 mm indentation depth. 29
8 Custom Python control interface, showing indentation and grid parameters, live force readings, the live camera feed, and the event log. 30
9 Indentation grid preview 31
10 Comparison of normal-force estimation performance across the four calibration architectures in terms of F_z mean absolute error (MAE). 40
11 Predicted versus ground-truth normal force for the three convolutional calibration architectures. The dashed line represents ideal prediction. 41
12 Contact-localisation predictions for the three convolutional calibration architectures. Ground-truth and predicted contact coordinates are shown for the respective held-out evaluation sets. 42
13 Predicted versus ground-truth indentation depth for the Physics-Informed Two-Stage Pipeline. The dashed line represents ideal prediction. 43
14 Representative output of the Physics-Informed Two-Stage Pipeline on a single held-out frame. The fringe crop (green) and mirror crop (blue) used for depth and force estimation and contact-location estimation, respectively, are annotated on the source frame. Predicted and ground-truth depth, contact position, and F_z are shown together with the corresponding prediction errors. 44

List of Tables

1 Comparison of representative vision-based tactile sensors 21
2 Training configuration for the four calibration architectures. 34
3 Calibration performance on held-out test sets 39
4 Comparison of the Physics-Informed Two-Stage Pipeline with representative tactile systems 46

Nomenclature

Symbols

Symbol	Description
F_x	Force component in the x-direction
F_y	Force component in the y-direction
F_z	Normal force component in the z-direction
F_{mag}	Magnitude of the resultant force
x, y	Contact coordinates on the sensing surface
d	Indentation depth or photoelastic membrane thickness, depending on context
$\hat{y}$	Predicted value of quantity y
$\sigma_x, \sigma_y, \sigma_z$	Principal stress components
n_0	Refractive index of the unstressed material
n_x, n_y	Principal refractive indices under stress
c_1, c_2	Stress-optic coefficients
C	Relative stress-optic coefficient, $C = c_1 - c_2$
λ	Wavelength of illuminating light
θ	Angular quantity associated with the optical or stress configuration
R^2	Coefficient of determination

Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
CNN	Convolutional Neural Network
ICP	Iterative Closest Point
MAE	Mean Absolute Error
MLP	Multi-Layer Perceptron
RGB	Red, Green and Blue
RMSE	Root Mean Square Error
ToF	Time of Flight

1 Introduction

One of the most difficult problems in modern robotics is achieving human-level dexterity. While computer vision is already mature, enabling high-fidelity object classification from remote measurements, tactile perception is not yet at this level of maturity. This difference arises from the inherent distinction between the two modalities: vision is generally passive. It operates at a distance, whereas tactile perception is contact-based and depends on physical interaction with the environment. To directly measure the mechanical properties of an object through touch, a robotic system requires a compliant interface that can accommodate irregular geometry while remaining sufficiently robust to transduce forces in multiple directions.

In addition to the hardware, tactile sensing is necessary to close the control loop during object manipulation. When operating in unstructured environments and with deformable objects, a manipulation controller can be especially challenged by the lack of tactile information about the location and magnitude of contact forces. To overcome this, tactile sensing has progressed from traditional capacitive array sensors to high-resolution optical skins that can generate topographic maps of an object's surface and determine contact conditions in much finer detail.

With the high pixel resolution of modern cameras, vision-based tactile sensors have become one of the most popular solutions. One example is the GelSight sensor [1], a soft elastomer with a micron-scale reflective coating used to measure a surface's geometry. These geometry-based devices can provide detailed information about an object's shape. Still, they may fail to separate complex force vectors from surface deformation, necessitating advanced inference to map observed deformation to applied forces.

A useful alternative is photoelasticity, which allows the direct optical visualisation of internal stress distributions. By using the birefringence of transparent polymers subjected to load, these sensors form isochromatic fringe patterns which carry information about the stress state within the sensing material. Recent advances have shown that deep learning models can capture these intricate patterns and simultaneously predict both normal force and contact location. However, integrating these sensors into compact robotic digits remains an engineering challenge. The current photoelastic finger designs all take the form of an optical stack within the finger. This can lead to non-uniform illumination over the sensing membrane.

However, a drawback of top-down illumination is that the light intensity at the edges of

the sensing membrane may be considerably lower than at the centre [2]. This results in potential blind spots and less sensor reliability in edge contacts and pinching manoeuvres. Such restrictions are especially important for manipulators with a high degree of freedom that interact with the object at multiple points, such as across its entire surface. Such illumination problems and photoelastic sensing in general have been studied so far mostly at the fingertip level. Although photoelastic tactile sensing has been demonstrated at the fingertip scale, its extension to the larger, less constrained sensing area of a robotic palm remains underexplored.

In this work, a photoelastic tactile sensor for robotic palms is proposed. A stack of multiple diffuser sheets and an intensity-reducing grid illuminate the membrane from the back, spreading the light more evenly over the sensing area. The fringe pattern is observed using a source-side polariser and a camera-side polariser oriented at 90° to it, forming a crossed-polariser configuration. The optical path folding is realised with an angled mirror, which reflects the bottom surface of the membrane towards a camera located off-axis but not below the sensor, keeping the package low-profile and appropriate for palm-scale sensing. A prototype is described and evaluated primarily for normal-force estimation and contact localisation, with recorded lateral-force components additionally examined, which serve as a foundation for extending high-resolution optical tactile sensing to dexterous robotic hands. This work also advances the general idea of optical and vision-based tactile sensing, in which tactile information is recovered from cameras rather than dense electronic readout elements, thereby enabling measurements in the presence of electromagnetic interference and providing high-resolution measurements with relatively simple imaging interfaces.

2 Background and State-of-the-Art

As robotic manipulation moves from structured industrial environments into less structured environments, tactile sensing can provide complementary information to vision, particularly when contact areas are obscured or there is geometric uncertainty [1, 3]. Local interaction forces can be obtained by using conventional tactile arrays or point-contact fingertip sensors. However, they may be limited in their use for whole-hand grasping and palm-scale manipulation due to their sensing area and spatial resolution. In contrast, vision-based tactile sensing has been developed to overcome these limitations by observing the deformation of a compliant sensing medium from the internal camera to obtain a high-dimensional optical representation of contact, without relying on dense electronic components to be placed at the contact surface [1, 4]. This method has since been extended from surface deformation and topography to the visualisation of internal stress using photoelastic materials. However, deep learning has begun to be applied to mapping from optical observations to force, contact position, and object geometry.

2.1 Vision-Based Tactile Sensing for Robotic Manipulation

The basic concept of tactile sensing is to use an internal camera to observe a compliant contact medium. By decoupling the mechanical sensing interface from the electronic transducer, image pixels can be used to characterise contact deformation, providing more information than discrete sensing elements such as [1, 4]. A representative example is GelSight, an elastomeric material with an opaque reflective membrane used to capture surface deformation when in contact with an object [1]. Multidirectional LED illumination and photometric stereo can reconstruct the surface geometry with the reported spatial resolution of around 20–30 μm [1]. In addition, embedded marker patterns can be used to measure lateral displacement, thereby enabling estimation of shear deformation and detection of slip [1].

Highly compliant sensing media provide an alternative for applications involving larger deformations. The Soft-bubble sensor replaces a solid elastomeric block with an inflated latex membrane approximately 0.4 mm thick, observed using an off-the-shelf time-of-flight (ToF) depth camera (PMD pico flexx) [3]. The membrane forms an inflated spherical surface approximately 50 mm in height and undergoes substantial deformation during contact, producing a dense 3D representation of contact geometry [3]. The resulting point cloud has been used for object classification with convolutional neural networks, including ResNet-18, as well as object pose estimation and tracking using the Iterative

Closest Point (ICP) algorithm [3].

Topography-based tactile sensors can obtain detailed surface deformation; however, the relationship between displacement and applied load must be determined to estimate the force. Such a relationship can be complicated in non-uniform contact and large deformation, especially when the sensing medium has nonlinear or viscoelastic characteristics in the presence of non-uniform contact and large deformation, as described in [1, 3]. Learning-based methods can then emulate this mapping between the observed deformation and the applied force without needing to build an entire analytical model.

2.2 Large-Area and Palm-Scale Vision-Based Tactile Sensors

Most optical tactile sensors have been designed as small fingertip sensors, whereas for whole-hand manipulation, tactile information is needed over a larger, potentially curved surface. The introduction of optical sensing into palm-scale interfaces imposes new mechanical and optical constraints on the sensing system, including larger deformations and tighter optical path and illumination distributions [5, 6].

Liu and Adelson designed and over-moulded a passively bendable 3D-printed cantilever beam with a soft silicone gel to achieve both structural and material compliance. The compliant structure also enables the sensing surface to conform to grasped objects, thereby increasing the contact area compared with more rigid structures [5]. The deformable structure is filled with flexible silicone-encased LED filaments that illuminate it, and embedded cameras record the resulting tactile imprints [5]. This shows how it is possible to embed optical tactile sensing in soft palm structures without adding hard parts into the deformable part.

There has been some research on curved tactile interfaces as well with binocular vision. GelStereo Palm uses a binocular camera system to create 3D tactile point clouds atop a hemispherical silicone layer [7]. The distortions introduced by the refraction through the air, acrylic support and silicone are currently not taken into account in conventional stereo reconstruction. The Refractive Stereo Ray Tracing (GP-RSRT) model mitigates these effects and demonstrates an average error in 3D reconstruction of 0.21 mm, as shown in [7]. The sensing volume was further expanded by using a solid silicone hemisphere of 65 mm diameter in GelStereo Palm 2.0, such that a contact depth of about 10 mm can be achieved [8]. An active shape exploration framework was then developed based on Gaussian Processes to guide sensing toward areas of high uncertainty in the object's geometric shape [8].

Another way to achieve large-area tactile sensing is by combining photometric stereo and structured illumination to estimate forces over a 3D conical shell with an area of roughly 4,800 mm^2, as developed by Insight [6]. The sensor is a rigid aluminium internal structure overmoulded with a flexible elastomer and housing a camera. The structured-light patterns produced by a custom collimator carry information about the deformation over the surface [6]. The approaches show that optical tactile sensing is not restricted to fingertip shapes, but that camera placement and illumination may be relevant when incorporating such sensors into small robotic hands. Cameras can be placed directly behind the sensing medium in the internal volume that is not needed for actuation, transmission, or even structural purposes [5, 6].

2.3 Photoelasticity-Based Tactile Sensing

An alternative way to measure surface topography is to examine the stress in a clear, photoelastic material. A photoelastic polymer is used to generate birefringence that alters the propagation of polarised light under mechanical loading. When observed with an appropriate polariscope configuration, the resulting phase difference between the principal stress directions produces visible isochromatic interference fringes, as reported by [9] and [10]. The information in these fringe patterns is related to the stress distribution in the sensing medium.

In early photoelastic tactile sensors, discrete optical components were employed to detect the changes in transmitted or reflected light. They created a dynamic tactile sensor by combining a transmission polariscope, optical fibres, an LED emitter, and a PIN photodiode (PH) [9]. The sensor was capable of providing a continuous signal up to 6N normal force and also contained a dynamic signal to detect slip, reported to detect slip velocities as low as 0.1 mm/s [9]. Mitsuzuka et al. then experimented on the application of photoelastic polyurethane to measure the normal and tangential forces by means of optical reflection and photodiode-based readout methods [11]. In the case of photodiode-based sensing, the optical response is integrated over the sensing area; thus, spatial information for contact localisation and shape identification is limited, as reported by [9, 11].

The camera-based photoelastic sensing can recover the spatial information of the fringe pattern. The PhotoElasticFinger consists of a soft photoelastic silicone layer (Sorta Clear 12 with the elastic modulus of about 0.4 MPa) placed between two orthogonal polarising filters [10]. Light intensity is transmitted into an internal RGB camera and reflected to the RGB camera depending on the applied normal load. Viscoelastic hysteresis was addressed by a piecewise polynomial calibration, which allowed for estimating the normal

force in the range 0.5–8 N with an RMSE of less than 2.5 N in the dynamic grasping experiments [10].

Additionally, [2] showed that photoelasticity could be used with a camera to estimate force, contact position, and contact shape simultaneously. The system is of the reflection type, in which light passes through a polariser, travels through the photoelastic element, reflects from a mirror, and passes back through an analyser. The investigation included two sensor configurations: A layered and an integrated photoelastic gel configuration [2]. The study also highlighted the role of image quality and optical focus, as image sharpness significantly affected model performance [2]. Despite these advances, challenges remain in making photoelastic sensing across various surfaces available at the palm scale, including illumination uniformity and optical packaging. Illumination from the periphery or top down may result in a non-uniform intensity across a larger sensing surface, which can decrease fringe consistency at the edges [10, 2].

2.4 Deep Learning and Calibration Pipelines

Analytical models, data-driven regression or a mixture of both can be used to convert optical tactile observations into physical quantities. If the sensing medium exhibits non-uniform deformation, large deformations, or a viscoelastic response, analytical solutions become complicated. With this in mind, newer approaches, such as convolutional neural networks (CNNs) and other learning-based methods, are increasingly used to map camera observations to tactile quantities [6, 4].

Kakani et al. showed vision-tactile transfer learning with a compact stereo camera with a 10 mm baseline to observe a marker-patterned elastomer [4]. A VGG-16 network pre-trained in ImageNet was further trained to estimate 3D contact coordinates (X, Y, Z), rotational orientation, and normal force. The system was able to estimate the force with an average error of 0.022 N and the position and area of contact [4]. This illustrates how well pretrained convolutional architectures can learn to map from high-dimensional optical representation to multiple tactile quantities.

Direct application of deep learning to the photoelastic fringe images has also been done. Serrien et al [2] used a multi-task CNN to estimate the magnitude of the normal force, 2D contact coordinates, and the shape of the indenter, from a single photoelastic image. A variant of ResNet-18 was trained with a joint regression and classification loss and was evaluated with a normal-force RMSE of 0.145 N and a localisation RMSE of 0.446 mm, according to [2]. The paper also explored differential image representations, in which the

reference image with no load is subtracted from the loaded image,

$$I_{\Delta} = I(d) - I(0), \tag{1}$$

to minimise static background information and emphasise loading-induced changes, [2] introduced an "image weighting function. In other vision-based tactile systems, differential representations are also used to isolate deformation or displacement relative to a reference state [1, 4].

Learning-based models can predict spatially distributed quantities in tactile systems with larger or more complex contact regions. Insight leverages a fully convolutional ResNet architecture to project camera observations into a dense force representation of an object comprising about 3,800 surface nodes [6]. The Kuhn-Munkres algorithm maps image pixels to 3D surface coordinates, enabling the network to estimate force vectors across the sensor surface. The accuracy of the force magnitudes is about 0.03 N, and the force-direction error is approximately 5° [6]. These findings illustrate the benefit of deep learning in mapping high-dimensional tactile images to spatially distributed physical measurements.

Another way breaks the intermediate physical quantities apart from the final regression of force. In these staged architectures, image features are first fed to a regression model to estimate quantities such as indentation depth and contact position, and then to a second regression model to estimate force. This allows an explicit relationship to be established between the optical observation, the intermediate deformation state, and the final estimate of the force, and for each of these prediction components to be estimated individually.

2.5 Research Gap and Motivation for a Photoelastic Tactile Palm

Complementary methods to vision-based tactile sensing have been shown in the literature. Sensors with surface topology information like GelSight and Soft-bubble use the elastomer's deformation or depth measurement to obtain geometric information [1, 3]. These ideas have been extended to larger sensing surfaces using compliant structures [5], binocular vision, or structured illumination [8, 6]. In contrast, photoelastic sensors encode the internal stress through the change in birefringence of the sensor material, resulting in

an optical representation related to the mechanical state of the sensor material [9, 10, 2]. These optical representations can then be mapped to force, position, and other tactile quantities through learning-based calibration without requiring a complete analytical description of the sensing medium.

This comparison highlights differences between the behaviour of large-area tactile interfaces and camera-based photoelastic sensing. Most of the existing photoelastic systems have been developed to sense and estimate forces at the fingertip scale, and have demonstrated force or spatial estimation by implementing small optical arrangements [10, 2]. On the other hand, large-area and palm-sized systems have shown geometric deformation or contact-area sensing by means of either depth cameras or compliant structures or binocular imaging or structured illumination [3, 5, 8, 6]. These methods demonstrate that large-area optical tactile sensing is possible, but they employ different transduction mechanisms and measure various physical quantities via optical measurements.

Spatially resolved camera imaging and a large photoelastic sensing surface thus offer a potential way to gain both stress-related and spatial information across a palm-scale interface. Effective illumination uniformity, optical packaging, fringe patterns, and the relationship with the applied mechanical state must be taken into account in such a system. This optical measurement can then be used in a learning-based calibration pipeline to estimate contact position and force.

The proposed method integrates a photoelastic sensor into a palm-scale prototype. It uses a folded optical path and distributed back-illumination so that the entire optical path is contained within the small space of a hand-held prototype. The calibration framework separates deformation and localisation estimation from subsequent force regression, which can be performed in a structured way, from the photoelastic response to the desired tactile quantities.

Table 1: Comparison of representative vision-based tactile sensors

Sensor / Study	**Transduction Modality**	**Primary Tactile Outputs**	**Spatial Resolution / Surface Scope**	**Optical Path & Packaging**
GelSight [1]	Photometric stereo + marker tracking	Height map, 3D force, slip entropy	High (20–30 μm); flat/domed fingertip	Direct vertical camera stack
Soft-bubble [3]	Time-of-flight depth sensing	3D point cloud, ICP object pose	Dense point cloud; inflated dome (175.4 cm^2)	Bulky internal cavity (>100 mm depth)
ROMEO Palm [5]	Flexible RGB LEDs + spy cameras	Contact area coverage (paint test)	Visual contact area; dual-compliant palm	Internal cameras embedded in cantilevers
GelStereo Palm 2.0 [8]	Binocular stereo + RSRT ray tracing	3D tactile point clouds, contact saliency	Spatial res.: 1.5 mm; solid 65 mm hemisphere	Binocular camera beneath support plate
Insight [6]	Photometric stereo + structured light	3D directional force map, self-posture	Localisation: 0.4 mm; conical shell (4,800 mm^2)	Single camera inside hollow skeleton
Kakani et al. [4]	Stereo tracking + VGG-16 regression	3D force, $X/Y/Z$ position, contact area	X/Y error ≈ 1 mm; hemispherical fingertip	Internal stereo camera (10 mm baseline)
PhotoElastic-Finger [10]	Transmission photoelasticity	Normal force magnitude (0.5–8 N)	Single-point (no contact localisation)	Direct vertical camera stack
Serrien [2]	Reflection photoelasticity + Multi-task CNN	Normal force (F_z), (x, y) position, shape	Localisation: 0.446 mm; flat disc	Direct vertical camera stack

3 Principles of Photoelastic Tactile Sensing

When a transparent polymer is subjected to mechanical stress, it has a different refractive index along different directions. When light is transmitted through a stressed polymer material, it travels at different speeds depending on the plane of polarisation, which is polarised relative to the principal stress axes. This stress-induced birefringence is the physical phenomenon exploited in the photoelasticity of the finger [10]. It enables a transparent sensing membrane to show the mechanical state across the entire surface rather than at a few discrete measurement points.

$$I_o = I_i \sin^2(2\theta) \sin^2\left(\frac{\pi d(\sigma_1 - \sigma_2)C}{\lambda}\right) \tag{2}$$

The stress-optic law captures the relationship between stress and the resulting change in refractive index. In its general tensor form, the change in the refractive index tensor is proportional to the stress tensor [10]:

$$\Delta n_{ij} = -c_{ijkl}\sigma_{kl}, \tag{3}$$

where c_{ijkl} is the stress-optic tensor. For a thin membrane under plane stress, symmetry reduces this to two independent stress-optic coefficients, c_1 and c_2, giving the principal refractive indices as [10]:

$$n_x = n_0 - c_1\sigma_x - c_2(\sigma_y + \sigma_z), \quad n_y = n_0 - c_1\sigma_y - c_2(\sigma_x + \sigma_z), \tag{4}$$

where σ_x, σ_y, σ_z, and n_0 represent the principal stresses and the refractive index of the unstressed material. The quantity of practical interest is the difference $n_x - n_y$, since it governs the relative retardation, or angular phase shift, between the two polarisation components as they pass through a membrane of thickness d:

$$\Delta_{xy} = \frac{2\pi dC}{\lambda}(\sigma_x - \sigma_y), \tag{5}$$

where $C = c_1 - c_2$ is the relative stress-optic coefficient and λ is the wavelength of the illuminating light.

This phase shift cannot be measured directly but only indirectly as a change in intensity by surrounding the birefringent membrane with two polarisers to create a polariscope. The same phenomenon is described by the photoelastic intensity relation written in Eq. (2), and the angle θ between the principal stress direction and the polariser axis is also considered as a part of the phase term to obtain the light intensity emerging from the analyser [9]. This expression is a special case of the derivation in Eq. (5), for a thin, uniformly loaded layer. Two different ideas have been used to use this phase shift for a useful signal. Here, the birefringent membrane is considered as a transmission element between crossed polarisers, so that the transmitted light undergoes a phase shift that is partially transmitted through the analyser. The intensity obtained is recorded across the entire membrane as a spatial fringe pattern, where each point carries information about the local difference and direction of the principal stresses [9]. The second method is the birefringence light scattering method, which was used by Mitsuzuka et al. The sensor does not measure the intensity of light transmitted between crossed polarisers, but light scattered out of the light path into photodiodes that can detect the scattered light [11]. A multi-axis force measurement without a polariser stack can be achieved using this scattering approach. It is less suited, however, to high-spatial-resolution sensing, as the signal is averaged over the static spatial region associated with each photodiode and does not capture surface stress variation across that region.

This distinction becomes especially pronounced when transitioning from a single digit to the broader surface of a robotic palm. In a readout by scattering, the number of photodiodes per unit area would need to increase with sensor area to maintain spatial information. In contrast, a transmitted-light fringe image can be captured using a single camera. It can capture the entire stress field, irrespective of its size or irregularity. The sensor used in this work is thus based on the principle of cross-polarised transmission. One polariser is installed on the source side of the photoelastic membrane, and a second polariser is placed in front of the camera, with its polarisation axis at 90° to the source-side polariser. Using the crossed-polariser setup, the phase shift due to the stress in the membrane, as described by Eq. (5), creates a fringe pattern of transmitted intensity across the membrane, as described by Eq. (2) across the palm-scale-sized membrane.

4 Sensor Design and Fabrication

This section explains the rationale behind the sensor's optical path and sensing membrane, including the approaches used to achieve uniform illumination and the material selected for the sensing layer.

4.1 Material Selection

Three materials were evaluated as potential sensing membranes: Solaris, Sylgard 184 and a two-part casting resin. The same illumination and crossed-polariser configuration were used to load samples of each material with the same thickness and size. Differences in the resulting fringe patterns could therefore be attributed primarily to differences between the candidate materials. The fringe pattern formed under a certain stress depends on the material properties and optical configuration, as indicated in Eq. (5) by the stress-optic coefficient C of the material [10]. The higher the stress-optic coefficient, C, the denser the fringe pattern for a given thickness, wavelength and applied load. A direct qualitative comparison of the effective stress-optic response of the candidate materials is thus possible when comparing fringe density under the same loading [10].

For all comparisons, the highest number of fringes was observed for the Solaris material, suggesting the most effective photoelastic response under the tested loading conditions and making it the most sensitive material to the applied loading conditions. This is especially true for a sensor mounted on a palm, since the grasping forces exerted on the sensor are expected to be much lower than at a dedicated fingertip. A material with a low applied force that still produces a measurable phase shift will enable the use of thinner membranes.

There were also fringe patterns visible on the casting resin, but it was not suitable. After curing, it was sufficiently stiff that it could not conform to irregularly grasped objects. This stiffness was compounded by the palm's broad contact requirements, which necessitate compliance over a large area without mechanical constraint, enabling stable contact across the palm. Therefore, casting resin was not used, even though it exhibited a photoelastic response. The fringe pattern obtained with Sylgard 184 under similar conditions was not chosen as it was relatively sparse. Solaris was thus chosen as the material for the sensing membrane for the rest of this work.

4.2 Prototyping

The sensor was designed in two stages: the initial design shown in Fig. 1 and the final design shown in Fig. 2b. The first iteration used a configuration typical of optical tactile sensors at the scale of fingers: the camera was placed directly beneath the sensing elastomer, viewing it along the surface normal, and illumination was provided laterally via a cavity surrounding the membrane's perimeter. The entire assembly fit within a 3D-printed cylindrical chassis, with dimensions chosen to minimise the package size given the components' dimensions.

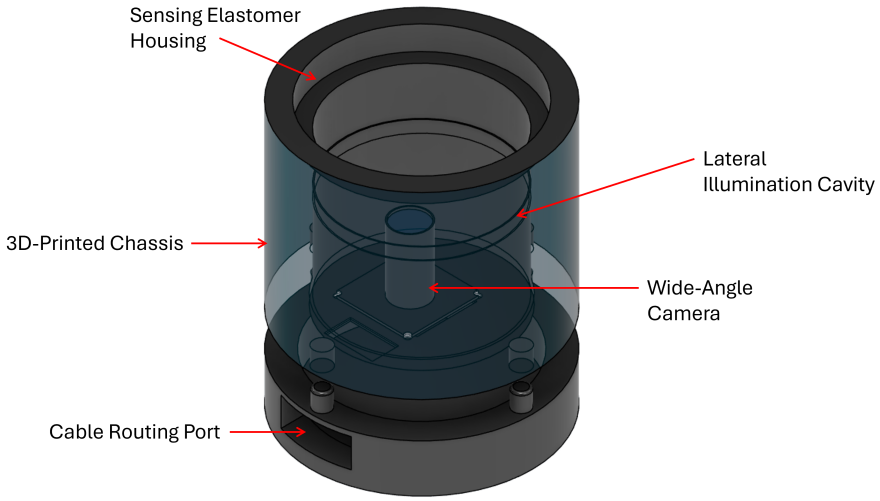


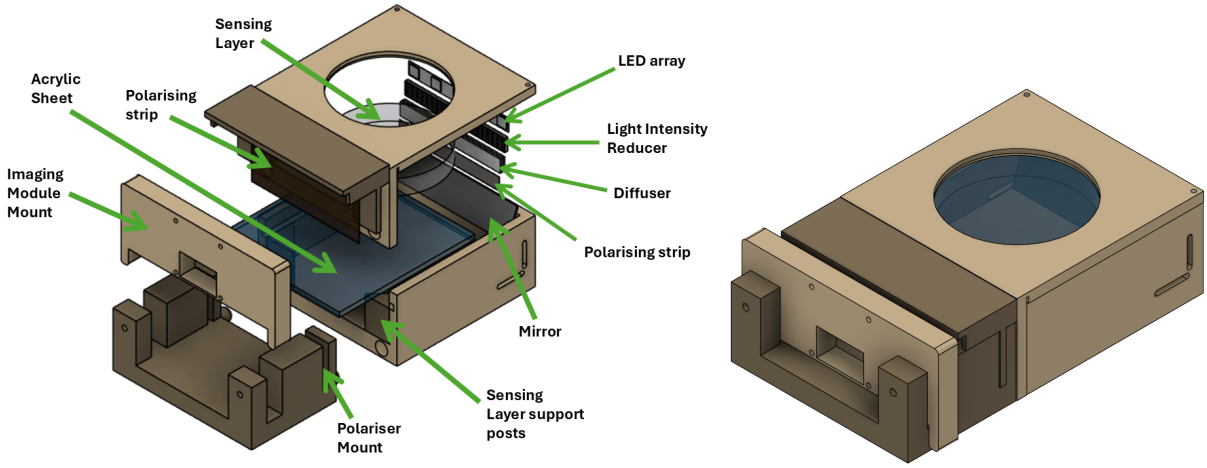
Figure 1: Initial prototype

This setup had two drawbacks. The contrast of the fringe pattern on the membrane was insufficient to ensure reliable readout, and the optical path was vertically stacked, which was ill-suited to the intended application. Placing a camera directly under the sensing surface leaves space beneath the palm for imaging hardware, where actuation and structural support are needed in a dexterous hand. The isolated compactness of a design is not necessarily the same as the compactness when it is embedded within a hand.

The second iteration then reversed the direction of the optical path, rather than shrinking it. The camera was then adjusted to view the sensing layer from the side, slightly below the membrane, and a mirror was added under the membrane, angled upward, to reflect the bottom of the sensing layer through the camera. This provides two complementary views using a single imaging device: the fringe field is observed across the cross-section of the sensing layer, and the mirror reflects the bottom of the membrane, yielding a second view from which contact points can be identified.

This is also the time the imaging device was rethought. The small footprint of a Raspberry

Pi camera module was selected for its suitability for the application, but was later replaced by a wide-angle USB camera attached to a laptop, as it was found insufficient for resolving fine fringe detail [1]. This means that the image resolution is significantly higher and more storage space is available for data collection, but at the expense of reduced compactness and a new data storage location (the laptop). The movement of the camera laterally also positions the imaging module in the area appropriate to the wrist of the hand that the camera is designed to work with, allowing the palm area of the hand to be used solely for the sensing membrane, as seen in the assembled prototype in Fig. 2b.



(a) Exploded view

(b) Assembled view

Figure 2: Final prototype: (a) exploded view showing the discrete illumination, polarisation, and mounting components; (b) assembled view.

The second iteration was not an afterthought, but instead a design requirement for modularity, as seen in the exploded view of the final design in Fig. 2a [12]. The illumination path is created by stacking discrete, independently mounted components: an LED array, a light-intensity-reducing grid, a diffuser stack, and a pre-polarising strip. This was to enable the optical components and polariser orientation to be changed without reconstruction of the assembly. The sensing layer is mounted on a set of support posts beneath a separate acrylic sheet, and the mirror is placed beneath it; both the imaging module mount and the polariser mount are detachable from the main housing. This allowed for the above material comparison to be made in practice as candidate membranes could be placed in an optical path other than the baseline path. The assembly is covered with a light-excluding lid to prevent ambient light from affecting the membrane, which would diminish the contrast of the fringe pattern.

4.3 Diffused Back-Illumination and Folded Optical Path

The flat, low-cost, and scalable light source was an off-the-shelf LED strip. The illumination was direct, creating high intensity on individual diodes and lower intensity between them, resulting in non-uniform illumination and potentially inducing artefacts similar to photoelastic fringes. To lower the intensity, a 3D-printed grid was placed in front of the strip, and three diffuser sheets were stacked to create a more uniform illumination field.

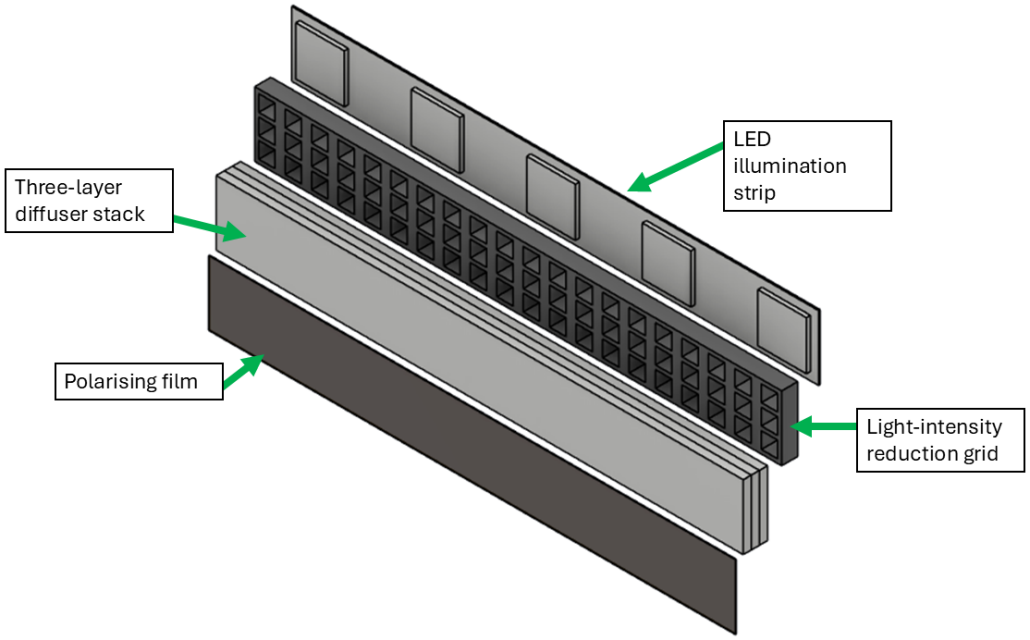


Figure 3: Exploded view of the back-illumination assembly

Figure 3 shows the layered back-illumination assembly used to make the illumination more uniform and reduce LED intensity across the sensing membrane.

To create a transmission-mode polariser, as described in Section 3, a single polarising sheet was positioned downstream of the diffuser stack, thus generating the source-side polariser. The LED assembly is placed behind the membrane, and a mirror angled relative to the membrane will reflect the backside of the assembly towards an off-axis camera. This optical path folding enables the camera to be positioned near the wrist rather than directly beneath the palm. The second polariser is in front of the camera, and is set at 90° to the source-side polariser. The unstressed light is not transmitted through this crossed-polariser configuration. Still, the light that the stress in the membrane has modified is transmitted, producing the intensity variations as photoelastic fringes.

5 Experimental Setup and Calibration

This section describes the automated rig used to collect labelled force, position and fringe-image data for sensor calibration.

5.1 Experimental Setup

A calibration dataset was needed that was systematically indented at known positions and depths and that simultaneously recorded fringe images and the forces measured in the ground truth. The manual approach of reproducing the contacts over the sensing area would have been time-consuming, so an automated indentation rig was built around a computer-controlled Cartesian stage (see Fig. 4). The sensor was securely placed on the stage bed, and the spherical indenter was mounted on the vertical axis and then moved in the x, y, and z directions. To calibrate the force measurements, a mini force/torque sensor was placed between the indenter and stage, providing "ground truth" values of the force applied to the probe [13].

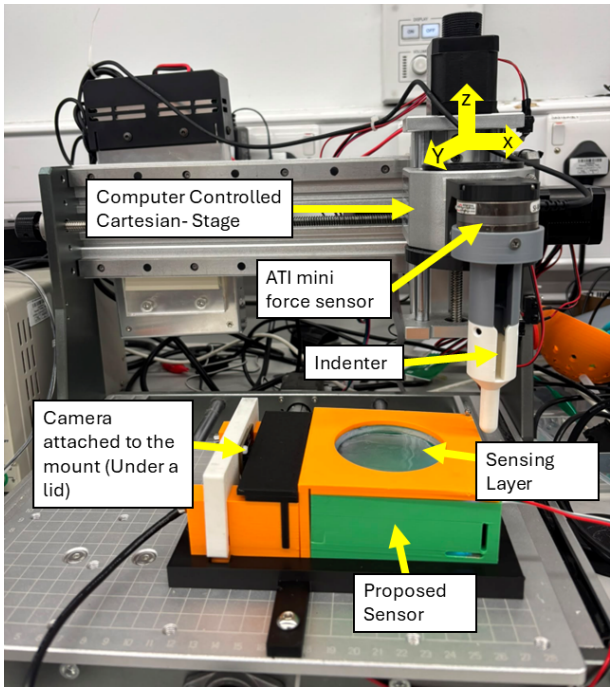
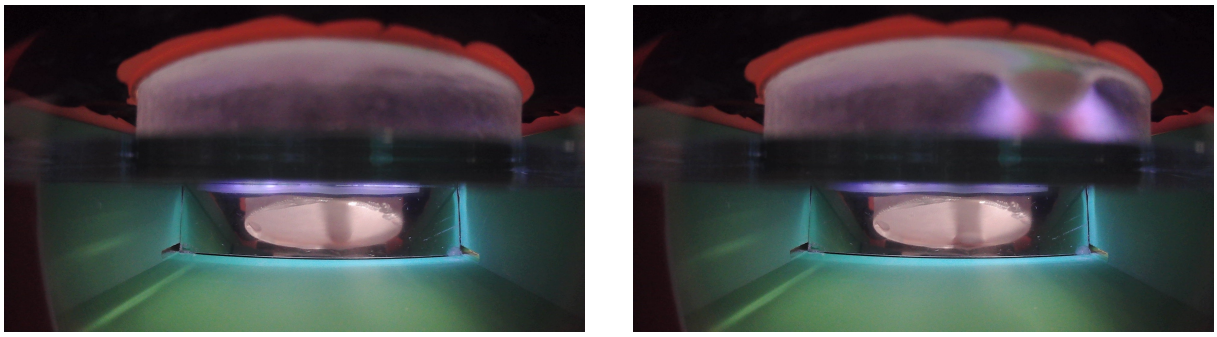


Figure 4: Sensor mounted on the computer-controlled Cartesian stage, with the ATI Mini force/torque sensor and indenter positioned above. The camera is secured to the white mount, viewing the sensor cross-section, with a lid enclosing the optical path to minimise ambient-light interference.

Figure 5 shows the camera view of the sensing membrane before and during contact, illustrating the change in the observed photoelastic fringe pattern under loading.

As shown in Figure 6, the most significant optical features that are observed during



(a) No-contact condition (b) Applied contact

Figure 5: Camera view of the sensing membrane under no-contact and applied-contact conditions.

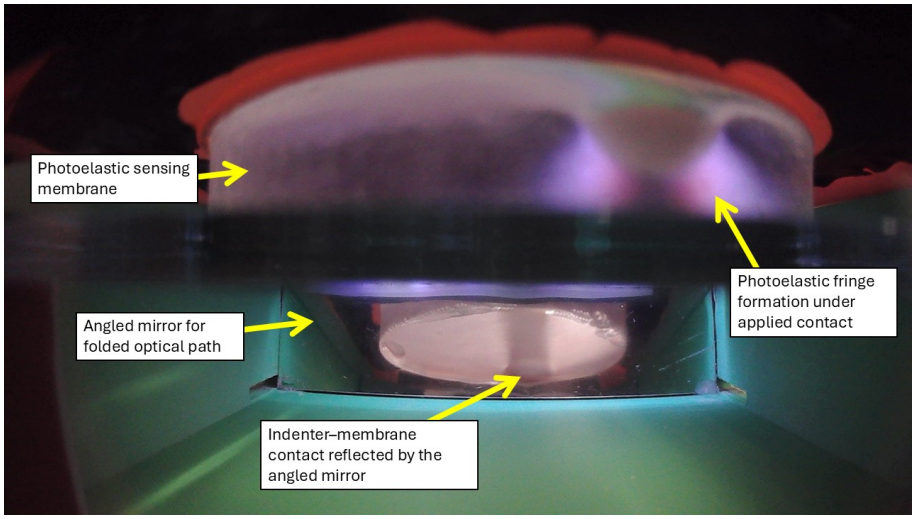
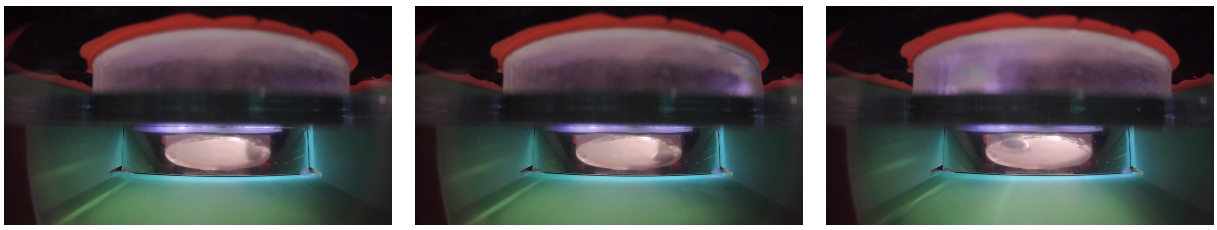


Figure 6: Annotated camera view under applied contact.



(a) Location 2 at 1 mm indentation. (b) Location 2 at 5 mm indentation. (c) Location 1 at 5 mm indentation.

Figure 7: Camera views illustrating changes in the photoelastic response with indentation depth and contact location. Panels (a) and (b) show the same contact location at 1 mm and 5 mm indentation depths, respectively, while panels (b) and (c) show different contact locations at the same 5 mm indentation depth.

indentation are the photoelastic fringe response in the sensing membrane, as well as the reflected contact region in the angled mirror.

It can be seen in Fig. 7 that the fringe pattern observed at a given contact position changes with indentation depth, while the fringe pattern observed at a given indentation depth changes with the contact position.

The motion of the stage and the acquisition of forces and images were controlled via a custom Python interface, as shown in Fig. 8. The interface set up the indentation grid, contact threshold, axis feed rates, force-sensor acquisition and calibration and output storage. It also featured real-time force measurements, the camera view, and the event log for monitoring and interrupting the automated sequence as needed.

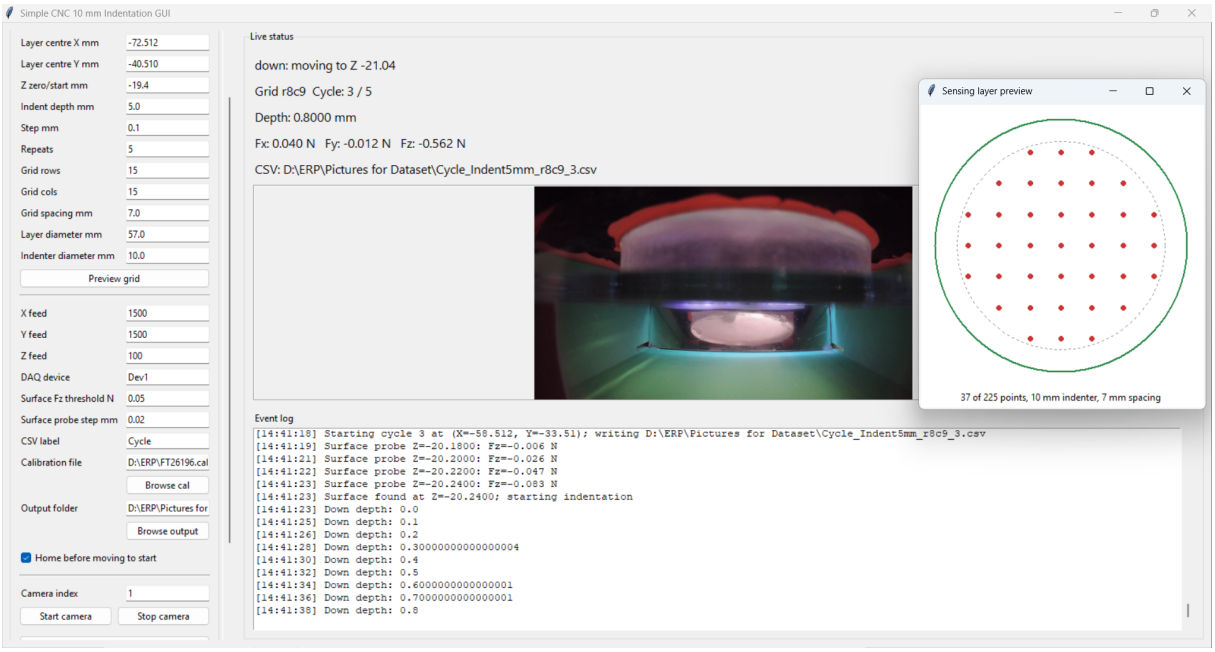


Figure 8: Custom Python control interface, showing indentation and grid parameters, live force readings, the live camera feed, and the event log.

For an automated run to start, the stage coordinate frame needed to be correlated to the physical sensor. The x and y positions of the centre of the sensing membrane in the stage coordinate system were determined empirically and used as the layer centre. A safe starting height on the z axis (the lowest point the indenter would touch the membrane if it were not stoppable, as shown below) was also determined manually and used as the z zero. All subsequent automated movements during the run were based on these manually set reference points, ensuring the correct positioning of the generated grid, regardless of the sensor's actual placement on the stage [6, 13].

Once the reference frame was determined, the indentation grid was automatically generated based on four parameters: number of rows and columns of the grid, the spacing between the grid points, the diameter of the sensing membrane, and the diameter of the indenter. Points that did not fall within the membrane diameter were automatically removed from the grid, and the indenter diameter was adjusted to avoid hitting the membrane edge as the indenter descended into the membrane during the test. With these parameters, the resulting grid is shown in Fig. 9, a 15×15 candidate grid, but only 37

valid indentation points remain due to the 57 mm diameter of the membrane and the 10 mm diameter of the indenter.

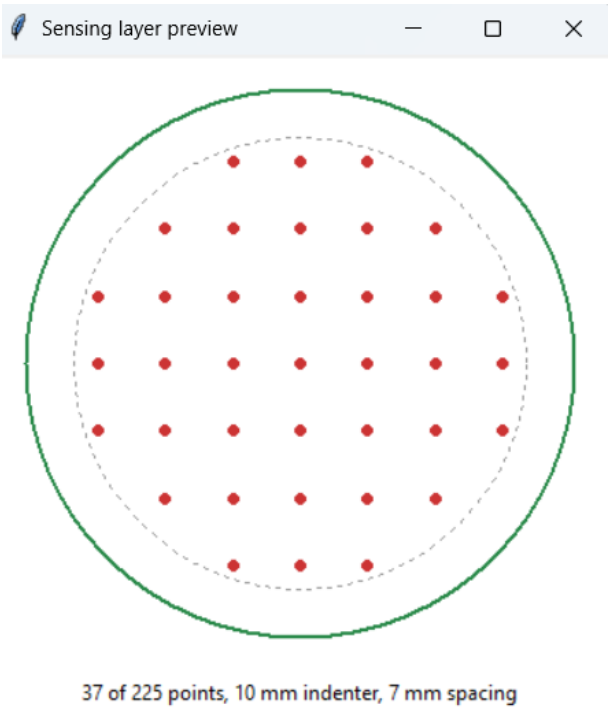


Figure 9: Indentation grid preview

The start of the membrane surface is crucial at these points; being at the correct height on the surface is essential, as deviations in its position would be interpreted as changes in the depth of indentation. The indenter was thus indented at 0.02 mm increments before every indentation and the z-axis force measured continuously. When this force was greater than 0.05 N, a threshold value to ensure early detection of contact, contact was registered, but was still sufficiently above the sensor noise level. The pointing height was then measured and used as the true local zero for the point from which the indentation had started.

After the contact was found, the indenter was lowered at 0.1 mm increments until it reached a depth of 5 mm, or fifty increments [13]. The three components of the forces (F_x, F_y, F_z), the position of the indenter (x and y displacement from the centre of the membrane) and twenty-five frames of the resulting fringe pattern were recorded at each depth before the indenter was moved to the next increment. The load cell measures the compressive normal force in the raw acquisition stream, with the sign used as $F_z \leq 0$; all force magnitudes in the calibration and evaluation are absolute ($|F_z|$). Visual readout is kept in the negative sign convention. The depth-force profile is generated step by step rather than taking a single reading at the maximum depth, so the relationship between indentation depth, applied force, and the fringe pattern can be studied at each

intermediate step. After reaching full depth of 5 mm, the indenter pulled back the entire way before advancing onto the next grid point.

The 37 valid points were indented 5 times each, instead of once. Silicone-based photoelastic sensors are commonly used under repeated loading because the material that makes up the sensor is viscoelastic and exhibits measurable hysteresis between load cycles [10]. This cycle-to-cycle variation can thus be stored within the dataset, rather than assuming a single indentation cycle represents the entire dataset. Considering all this, 185 indentation cycles were applied to the sensing surface, and, in total, about 230,000 individual fringe images were produced over the fifty indentation steps and the twenty-five photographic frames per step. For each image, a synchronised three-axis force reading was taken, and a known contact position was extracted from the run's CSV log for that image, from the corresponding row.

5.2 Model Architecture and Training

Calibration is used to create a relationship between the images captured by the sensor and the contact position and contact force. In vision-based tactile sensing, various data-driven calibration methods have been developed to measure physical quantities from tactile images directly [4, 14]. Assume I is an image that is obtained from the sensor. The calibration problem is to learn a mapping.

$$\mathcal{C} : I \mapsto \mathbf{y}, \tag{6}$$

where $\mathbf{y}$ represents the physical quantities targeted by a given calibration pipeline. As no explicit analytical model directly relates the observed photoelastic fringe pattern to the corresponding force and contact position, the mapping is approximated using a neural network,

$$\hat{\mathbf{y}} = f_\theta(I), \tag{7}$$

where θ denotes the learnable network parameters. The networks were trained using the indentation dataset described in the preceding subsection. Four candidate architectures were implemented to investigate different relationships between image representation and target physical quantities.

Motivation Behind Each Representation

Different image representations were explored for each of the four architectures. The Joint Multi-Task and Force-Conditioned Networks keep RGB crops to investigate if colour information is available beyond intensity in fringe variations between neighbouring bands [15, 16]. The Physics-Informed Two-Stage Pipeline, on the other hand, employs single-channel greyscale images and tests whether the intensity and fringe structure are sufficient for calibration, thereby reducing the complexity of the input. The MLP converts the 2D crop into a 1D pixel-intensity array, which is used to determine if force can be estimated without spatial convolution.

For architectures using a differential image representation, the delta image is defined as

$$I_{\Delta}(d) = I(d) - I(0), \tag{8}$$

where $I(d)$ denotes the frame acquired at indentation depth d and $I(0)$ denotes the zero-indentation frame recorded at the same physical location. Two region-specific differential crops are used throughout this work,

$$I_{\Delta}^{F} = \mathrm{crop}_{\mathrm{fringe}}\left[I(d) - I(0)\right], \tag{9}$$

$$I_{\Delta}^{L} = \mathrm{crop}_{\mathrm{mirror}}\left[I(d) - I(0)\right], \tag{10}$$

referred to subsequently as the *fringe delta crop* and *mirror delta crop*, respectively [17].

The PyTorch CNNs were trained optimising with Adam [18] and with a maximum gradient norm of 5.0 in all other runs mentioned later [19]. The flattened-intensity model was a scikit-learn MLPRegressor fitted using the Adam solver. For all four architectures, location-disjoint partitions were used, in which samples were partitioned based on their physical contact locations, and these partitions appeared only during training or evaluation. All the training settings are summarised in Table 2. The Force-Conditioned network shared the 30-epoch default of the raw-crop run. The MLP used scikit-learn's `MLPRegressor` on flattened 1D pixel-intensity arrays, with its batch size set to the default `min(200, n_samples)` and its epochs set to `max_iter`. Early stopping was used for

both the MLP (patience of 20 iterations) and the Physics-Informed Two-Stage pipeline (patience of 15 epochs).

Table 2: Training configuration for the four calibration architectures.

Architecture	**Optimiser**	**Batch Size**	**Epochs**	**Split**
Joint Multi-Task (Raw)	Adam	32	30	Location-disjoint
Force-Conditioned (Differential)	Adam	32	30	Location-disjoint
MLP (1D Flattened Array)	Adam	200	2000	Location-disjoint
Physics-Informed Two-Stage	Adam	32	50	Location-disjoint

5.2.1 Joint Multi-Task Network with Raw Crop Inputs

The first model uses two pre-trained ResNet-18 models operating in parallel on separate raw image crops. The raw fringe crop, I_{fringe}, is fed into one branch, and the raw indenter-contact mirror crop, I_{mirror}, into the second branch. The feature embeddings for the corresponding features are

$$\phi_F = \text{ResNet18}_F\left(I_{\text{fringe}}\right), \qquad \phi_L = \text{ResNet18}_L\left(I_{\text{mirror}}\right), \tag{11}$$

which are concatenated to form the shared representation

$$\mathbf{z} = [\phi_F, \phi_L]. \tag{12}$$

The combined representation is provided to separate force and localisation heads,

$$\hat{\mathbf{F}} = h_F(\mathbf{z}), \qquad (\hat{x}, \hat{y}) = h_L(\mathbf{z}), \tag{13}$$

where

$$\hat{\mathbf{F}} = (\hat{F}_x, \hat{F}_y, \hat{F}_z). \tag{14}$$

The force and localisation heads are completely independent after the shared feature representation. Each head is trained together with the same ground-truth force and contact-position data [6].

The training objective is defined as an equally weighted sum of mean squared error losses,

$$\mathcal{L} = w_F \,\mathrm{MSE}\left(\mathbf{F}, \hat{\mathbf{F}}\right) + w_L \,\mathrm{MSE}\left((x, y), (\hat{x}, \hat{y})\right), \tag{15}$$

where $w_F = w_L = 1$.

5.2.2 Force-Conditioned Network with Differential Crop Inputs

The second architecture is the same as the first, but the input features are replaced by the differential crops corresponding to each input channel in Eq. (10). The fringe delta crop goes to the force branch, and the mirror delta crop goes to the localisation branch.

The force head predicts the normal force component,

$$\hat{F}_z = h_F(\mathbf{z}), \tag{16}$$

where $\mathbf{z}$ denotes the concatenated feature representation produced by the two ResNet-18 branches. The predicted force is subsequently incorporated into the localisation head,

$$(\hat{x}, \hat{y}) = h_L\left([\mathbf{z}, \mathrm{sg}(\hat{F}_z)]\right), \tag{17}$$

The stop-gradient operator $\mathrm{sg}(\cdot)$ is used as indicated. The estimated normal force thus influences the localisation prediction, whereas the force branch is independent of the localisation loss during backpropagation [20].

The force and localisation outputs are trained together using a multi-task loss, with the multi-task loss weights set equal and the force output corresponding to F_z. In contrast,

the localisation output corresponds to (x, y).

5.2.3 MLP Regression on Flattened 1D Delta-Crop Arrays

The third architecture is completely different from convolutional feature extraction, testing an intensity-based representation of the fringe delta crop rather than the spatial representation [21]. The fringe delta crop is flattened to 1D, resulting in a 1D vector of pixel intensities rather than preserving its 2D structure.

$$\mathbf{v} = \text{flatten}\left(I_{\Delta}^{F}\right) \in \mathbb{R}^{n}, \tag{18}$$

where n refers to the number of all intensity values of the crop without any convolutional or spatial processing done in advance. This vector is directly fed to a multi-layer perceptron regressor, called `MLPRegressor` from the scikit-learn library, and the regressor is trained to regress the three-axis force vector

$$\hat{\mathbf{F}} = \text{MLP}_{\rho}\left(\mathbf{v}\right), \tag{19}$$

where ρ represents the parameters of the MLP. This network predicts only the force, not the contact location, unlike previous network architectures. It lacks a convolutional structure; thus, its local spatial relationships and/or translation-equivariant feature extraction are not explicitly used; therefore, it serves as a baseline to measure the contribution of convolutional image processing.

The architecture was evaluated using a location-disjoint partition, as summarised in Table 2.

5.2.4 Physics-Informed Two-Stage Estimation Pipeline

The fourth one separates the estimation of indentation depth and contact location from the estimation of force. A DepthCNN is trained with a pre-trained ResNet-18 network to predict the indentation depth of a crop of the fringe delta, which is a single-channel greyscale image

$$\hat{d} = f_{\theta_d}\left(I_{\Delta}^{F}\right). \tag{20}$$

A second LocalisationCNN of the same general architecture predicts the contact position from the mirror delta crop,

$$(\hat{x}, \hat{y}) = f_{\theta_l}\left(I_{\Delta}^{L}\right). \tag{21}$$

Unlike DepthCNN, LocalisationCNN does not rely on depth data. The indentation depth and contact position are then concatenated and given to a ForceMLP

$$(\hat{F}_x, \hat{F}_y, \hat{F}_z, \hat{F}_{\text{mag}}) = f_{\theta_F}\left(\hat{d}, \hat{x}, \hat{y}\right), \tag{22}$$

where $\hat{F}_{\text{mag}}$ represents the expected value of the magnitude of the force. The force is then estimated from the estimated indentation depth and contact position rather than from direct image representation [22].

Unlike the two RGB ResNet architectures, the DepthCNN and LocalisationCNN take single-channel greyscale image crops as input. The greyscale input enables the models to operate directly on the spatial intensity variations in the photoelastic fringe pattern and the mirror region.

The prediction components are supervised using the Huber loss. The overall training objective is

$$\mathcal{L} = w_d \ell_\delta\left(d, \hat{d}\right) + w_l \ell_\delta\left((x, y), (\hat{x}, \hat{y})\right) + w_F \ell_\delta\left(\mathbf{F}, \hat{\mathbf{F}}\right), \tag{23}$$

where ℓ_δ denotes the Huber (SmoothL1) loss and w_d, w_l, and w_F denote the respective loss weights.

The background appearance at any particular location within the grid is fairly stable from depth to depth and cycle to cycle, since the camera is stationary between measurements. The location-disjoint partitions should then prevent any samples from the evaluation location also appearing in the training partition, thereby minimising the chance that the localisation test results are due to multiple time measurements of the same physical location rather than predictions for new locations.

6 Results

6.1 Evaluation Metrics

Performance of the calibration models was assessed using metrics chosen to measure force estimation accuracy, indentation depth estimation, and contact location accuracy. The mean absolute error (MAE) was used to measure the average absolute prediction error, and the coefficient of determination (R^2) was used to measure the proportion of variance in the ground-truth measurements explained by each model [23].

For a set of N test samples, the MAE for a predicted quantity y is defined as

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \,, \tag{24}$$

where y_i and $\hat{y}_i$ denote the ground-truth and predicted values, respectively.

The coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{25}$$

where $\bar{y}$ is the mean of the true values. The lower the MAE, the less error in the prediction, and the closer the R^2 value to unity, the better agreement between the predicted and actual quantities [23].

To determine the contact location, the x_{mm} and y_{mm} coordinates were computed separately from the MAE values. The Euclidean contact-location error is defined as

$$e_{\mathrm{loc}} = \sqrt{(x_{mm} - \hat{x}_{mm})^2 + (y_{mm} - \hat{y}_{mm})^2}, \tag{26}$$

where (x_{mm}, y_{mm}) and $(\hat{x}_{mm}, \hat{y}_{mm})$ denote the ground-truth and predicted contact coordinates, respectively.

Table 3: Calibration performance on held-out test sets

Method	**F_z MAE (N)**	**2D Localisation MAE (mm)**
Joint Multi-Task (Raw)	0.443	2.321
Force-Conditioned (Differential)	0.244	3.205
MLP (1D Flattened Array)†	0.330	N/A
Physics-Informed Two-Stage	0.486	0.204

Note: Localisation MAE is calculated as the mean Euclidean distance between the predicted and ground-truth contact coordinates. The Joint Multi-Task and Force-Conditioned architectures operate on RGB image representations, whereas the MLP and Physics-Informed Two-Stage Pipeline operate on greyscale representations. All four architectures were evaluated using location-disjoint partitions. The model with the superscript † estimates force only and does not predict contact location.

6.2 Quantitative Results

The four calibration architectures were evaluated on their respective held-out test sets. The Joint Multi-Task Network with raw crop inputs and the Force-Conditioned Network with differential crop inputs both take RGB image crops as input and output the force and contact location. The MLP on a flattened 1D delta crop is only used to predict force from a greyscale fringe delta crop. The Physics-Informed Two-Stage Pipeline uses single-channel greyscale crops to predict the indentation depth, contact position, and contact force.

The results of the evaluation are summarised in Table 3. To evaluate all four architectures, each sample from a physical location included in the evaluation set was removed from the training set because the evaluation locations were partitioned in a location-disjoint manner and the group-aware partitions were used. This provides a uniform evaluation of the prediction at a fixed number of points in the observed data, away from the training set.

6.2.1 Force Prediction

The Joint Multi-Task Network (with raw crop inputs) obtained an F_z MAE of 0.443 N. Compared with the raw-crop architecture, the differential-representation then yielded the best F_z MAE for the evaluated architectures, at 0.244 N. The differential Force-Conditioned architecture achieved better normal-force prediction than the raw Joint Multi-Task architecture under the tested conditions. The decrease in error aligns with the differential image representation, which aims to remove static image elements while

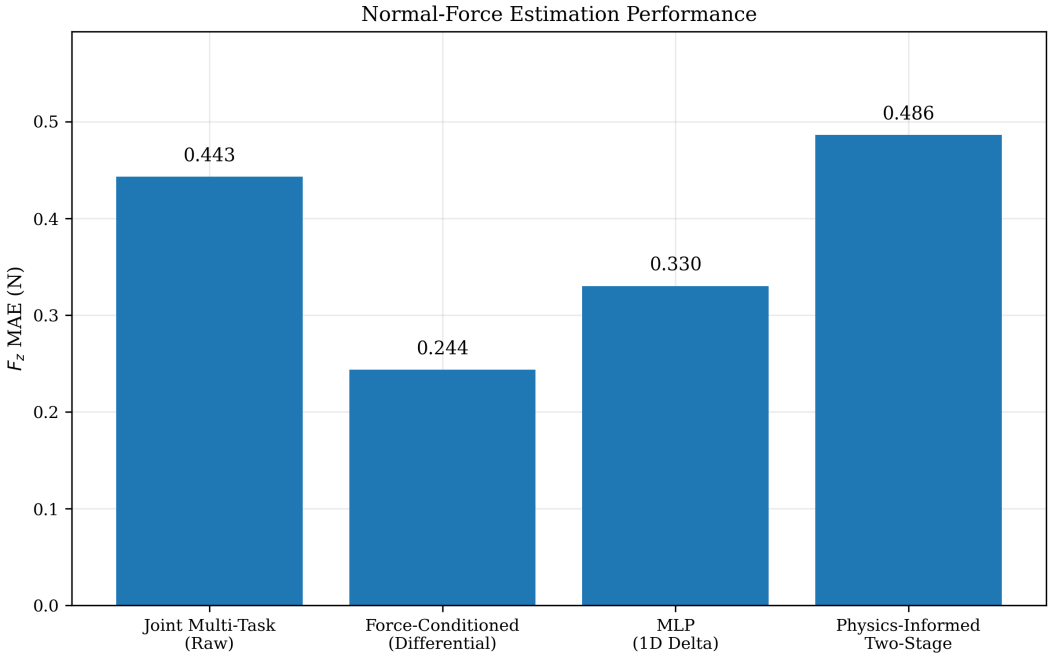


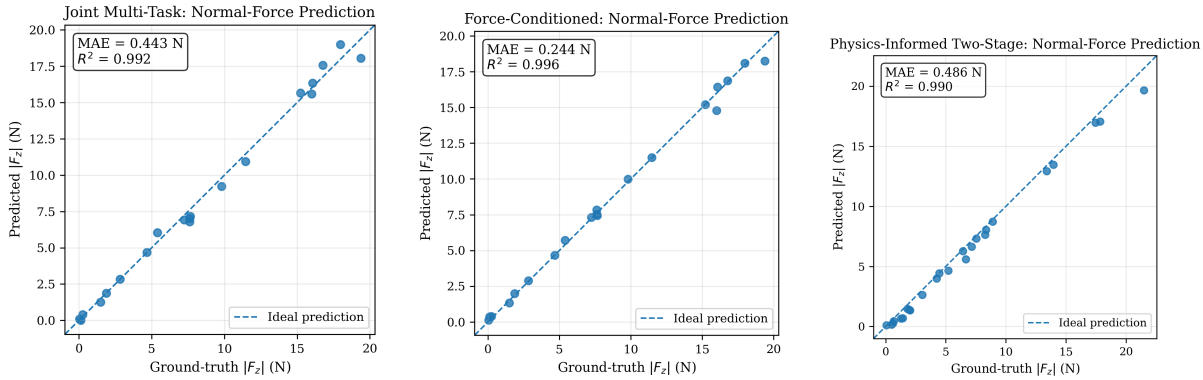
Figure 10: Comparison of normal-force estimation performance across the four calibration architectures in terms of F_z mean absolute error (MAE).

highlighting intensity changes resulting from indentation [15, 24].

The MLP with the flattened 1D delta-crop arrays performed with an F_z MAE of 0.330 N. Additionally, the model error for F_x and F_y was 0.036 N and 0.045 N, respectively. Despite completely removing all spatial and convolutional structure and keeping only the intensity basis, its F_z performance was competitive with end-to-end convolutional architectures. This outcome was also achieved with a location-disjoint evaluation protocol, in which the physical locations included in the evaluation set were not present in the training set, as in CNN architectures. The findings then show that the learned image representation contained sufficient information to enable force regression at previously unseen physical locations within the completed experimental campaign.

The Physics-Informed Two-Stage Pipeline performed with an F_z MAE of 0.486 N and an R^2 of 0.990. The model also achieved an F_x MAE of 0.035 N with an R^2 of 0.600, an F_y MAE of 0.037 N with an R^2 of 0.975, and a force-magnitude MAE of 0.457 N with an R^2 of 0.990. Within the incidental lateral-force variation present in the normal-indentation dataset, F_y, F_z, and F_{mag} showed high R^2 values. Results show that the lower R^2 was observed for F_x, suggesting greater variability in predicting the lateral component of the force, but the absolute MAE remained low.

The normal-force result is especially important since the two-stage architecture does not



(a) Joint Multi-Task. *(b) Force-Conditioned.* *(c) Physics-Informed Two-Stage.*

Figure 11: Predicted versus ground-truth normal force for the three convolutional calibration architectures. The dashed line represents ideal prediction.

directly estimate F_z from the image. Rather, the predicted indentation depth and contact position are used as intermediate representations to the ForceMLP. The resulting R^2 of 0.990 indicates that the learned intermediate quantities contain substantial information about the applied normal force.

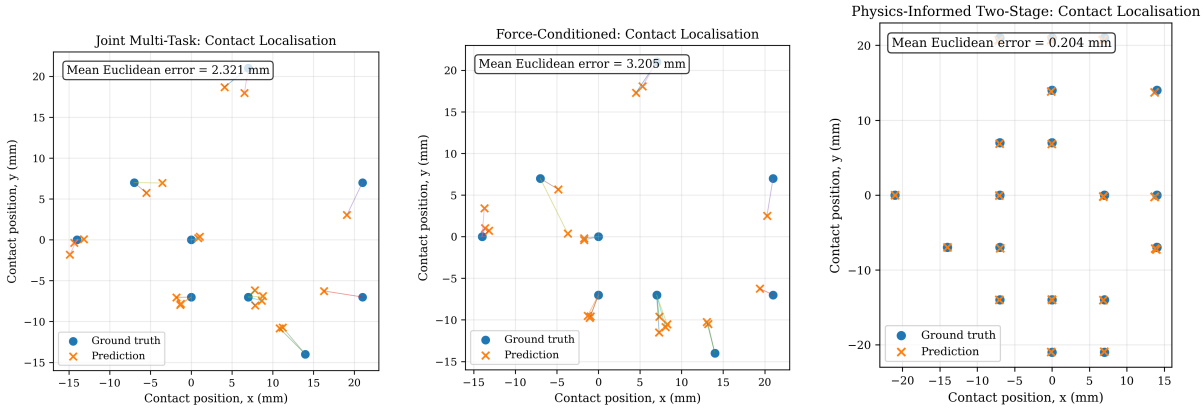
The Physics-Informed Two-Stage Pipeline exhibited the best depth and contact-location estimation, along with a good normal-force fit, and the differential RGB architecture achieved the lowest measured F_z MAE. The 1D flattened-array MLP was also found to be competitive for force prediction when using the common location-disjoint evaluation protocol.

6.2.2 Contact-Location Prediction

The Joint Multi-Task Network with raw crop input had an average error of 2.321 mm in terms of contact location. The Force-Conditioned Network with differential crop inputs had a higher mean error of 3.205 mm. The differential architecture reduced the x_{mm} MAE from 1.718 mm to 1.247 mm, while the corresponding y_{mm} MAE increased from 1.243 mm to 2.757 mm. Hence, the improvement along the X_{mm} axis was insufficient to offset the increase in the Y_{mm} error, resulting in an overall increase in the localisation error.

Since the MLP on flattened 1D delta-crop arrays does not estimate the contact location, it is not included in the localisation comparison.

The physics-informed two-stage pipeline made very accurate predictions of contact positions. The model achieved an x_{mm} MAE of 0.131 mm with an R^2 of 0.9997 and a



(a) Joint Multi-Task. *(b) Force-Conditioned.* *(c) Physics-Informed Two-Stage.*

Figure 12: Contact-localisation predictions for the three convolutional calibration architectures. Ground-truth and predicted contact coordinates are shown for the respective held-out evaluation sets.

y_{mm} MAE of 0.147 mm with an R^2 of 0.9998. The corresponding mean 2D Euclidean localisation error was 0.204mm. Based on these results, we can conclude that there is a very high degree of similarity between the predicted and measured contact coordinates in the evaluation set.

The dedicated LocalisationCNN in the two-stage architecture is consistent with high localisation accuracy. The localisation branch takes the mirror delta crop and estimates the contact position without relying on the force-estimation pathway. The differential representation also eliminates the static background appearance of the zero-indentation state, thereby enabling it to capture contact-associated changes in the mirror region [15, 24].

The two-stage pipeline was tested on a partition with a location-disjoint set defined by the physical grid-point identity. Thus, no samples collected at an evaluation location were included in the training set, including samples acquired at different indentation depths or across repeated cycles. This localisation performance is thus the generalisation to unseen, physical locations in the completed experimental dataset, without train-test leakage between repeated observations of the same grid points [25, 26].

However, the sub-millimetre coordinate errors and R^2 values close to unity indicate that the proposed sensor and two-stage calibration procedure accurately estimate the contact position across the sensing surface examined.

6.2.3 Comparison of Calibration Architectures

Depending on the two types of input representation and prediction architecture, the four calibration architectures show different performance. The Force-Conditioned Network achieved the lowest direct normal-force error using differential RGB inputs, which had the lowest measured F_z MAE of 0.244 N. The MLP with 1D delta-crop arrays, flattened with no spatial structure whatsoever, achieved a comparable F_z MAE of 0.330 N under identical location-disjoint evaluation criteria.

The Physics-Informed Two-Stage Pipeline enabled another calibration feature to jointly estimate the indentation depth, contact position, and contact force in two stages. Its DepthCNN achieved an MAE of 0.165 mm and an R^2 of 0.978, while the LocalisationCNN achieved an MAE of 0.131 mm and 0.147 mm for x_{mm} and y_{mm}, respectively. The R^2 values of 0.9997 and 0.9998 show that the predicted contact coordinates are highly consistent with the ground-truth measurements.

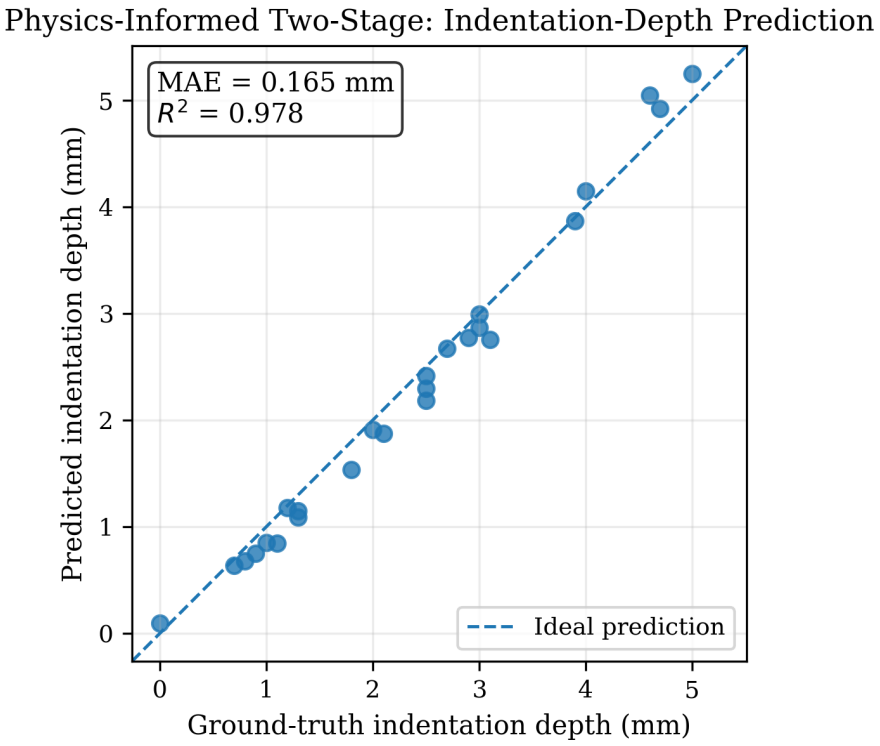


Figure 13: Predicted versus ground-truth indentation depth for the Physics-Informed Two-Stage Pipeline. The dashed line represents ideal prediction.

The two-stage pipeline had an F_z MAE of 0.486 N and an R^2 of 0.990 for force estimation. The two-stage pipeline provides estimates of indentation depth and contact location, although its absolute F_z MAE exceeds that of the differential RGB architecture. It also

uses these intermediate physical quantities to predict the force, rather than relying directly on image features, thereby establishing a more structured relationship between the sensing image and the final force prediction.

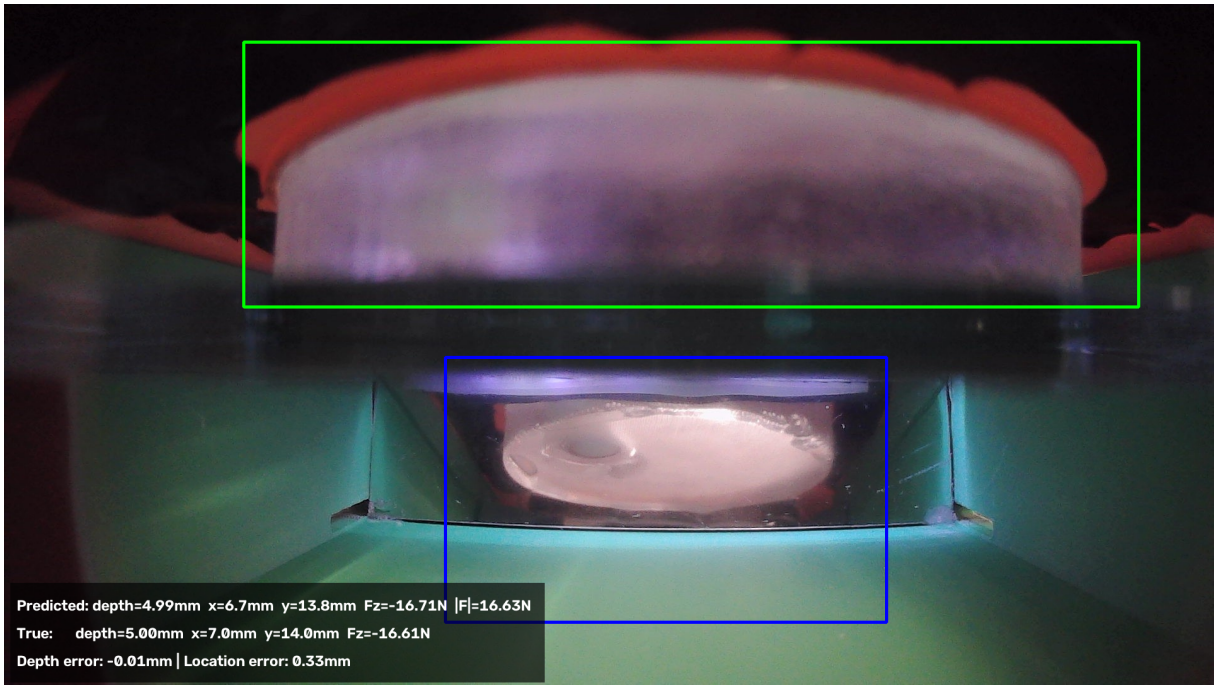


Figure 14: Representative output of the Physics-Informed Two-Stage Pipeline on a single held-out frame. The fringe crop (green) and mirror crop (blue) used for depth and force estimation and contact-location estimation, respectively, are annotated on the source frame. Predicted and ground-truth depth, contact position, and F_z are shown together with the corresponding prediction errors.

The comparison is also affected by the different representations of the image. The Joint Multi-Task Network and the Force-Conditioned Network take RGB images as input; the two-stage pipeline's DepthCNN and the LocalisationCNN take single-channel greyscale images as input, while the MLP has no spatial structure whatsoever but operates on a flattened 1D pixel-intensity array. Since colour representation, spatial dimensionality, and architecture vary together, the individual contribution of the image representation cannot be isolated from the architectural design using the present results.

The common location-disjoint protocol enhances architectural comparison by omitting one of the evaluation locations during training. The localisation errors are thus limited to prediction at unseen grid locations within the same experimental campaign [25, 26]. Combined, the Force-Conditioned architecture achieved the best normal-force prediction, while the Physics-Informed Two-Stage Pipeline provided the most accurate depth and contact localisation alongside normal-force estimation.

7 Comparison with State-of-the-Art

The proposed sensor is compared with the representative learning-based optical tactile systems for force estimation, contact localisation and spatial contact characterisation, such as Kakani et al. [4], Soft-bubble [3] and the ROMEO compliant tactile palm [5]. Any differences in how sensing was performed or in the evaluation process used are taken into account when interpreting the reported metrics.

Force Estimation

Kakani et al. [4] achieved an average force estimation error of 0.022 N in the range of 0.1–1.0 N. A photometric-stereo-based architecture was used to estimate the normal and shear forces, and the force errors were reported to be around 0.03 N by Insight [6].

The proposed Physics-Informed Two-Stage Pipeline obtained F_z MAE of 0.486 N, and R^2 of 0.990. The pipeline additionally achieved an F_x MAE of 0.035 N and an F_y MAE of 0.037 N, with corresponding R^2 values of 0.600 and 0.975. The force magnitude predictions were made with an MAE of 0.457 N, and an R^2 of 0.990.

The numerical force error is higher than that reported for the specialised optical systems taken into account in the present comparison. The proposed architecture, however, is implemented as a multi-stage tactile calibration process that also estimates the indentation depth and contact position on a palm-sized sensing surface. The system thus provides a wider range of tactile outputs within a single sensing architecture, rather than being optimised for force-regression accuracy.

Contact Localisation

Kakani et al. [4] claimed 1.396 mm accuracy on the X axis and 0.973 mm on the Y axis for the localisation of the contact. In the continuous working surface, the spatial error was reported to be about 0.4 mm by Insight [6].

The proposed Physics-Informed Two-Stage Pipeline (PI2SP) yielded an x_{mm} MAE of 0.131 mm, corresponding to an R^2 of 0.9997, and a y_{mm} MAE of 0.147 mm, corresponding to an R^2 of 0.9998. The results shown here demonstrate sub-millimetre accuracy of contact positioning over the evaluated sensing region.

This should, however, be compared with a certain caveat: in the context of different

evaluation procedures. This is a different spatial sampling and validation procedure than that used for the reported results from systems such as Insight and Kakani et al. in the present evaluation, which relies on discrete physical grid locations.

Table 4: Comparison of the Physics-Informed Two-Stage Pipeline with representative tactile systems

Sensor	**Force Accuracy**	**Localisation Accuracy**	**Sensing Principle / Architecture**	**Scope and Validation Modality**
Proposed Sensor	F_z: 0.486 N MAE	x: 0.131 mm MAE y: 0.147 mm MAE	Internal camera; CNN; Photoelastic layer	Planar palm surface; discrete physical grid evaluation
Insight [6]	F_n, F_s: ~ 0.03 N	~ 0.4 mm	Photometric stereo with structured illumination and deep neural network	3D conical shell; continuous active surface evaluation
Kakani et al. [4]	0.022 N average	X: 1.396 mm Y: 0.973 mm	Stereo camera; transfer-learned VGG16 regression	Flat elastomeric sensing surface; evaluation over 0.1–1.0 N
Soft-bubble [3]	*Not reported as MAE*	*Not reported as MAE*	Internal depth sensing and optical tracking; ICP-based pose estimation	Compliant membrane; evaluated through contact and geometric characterisation
ROMEO Palm [5]	*Not reported as MAE*	*Not reported as MAE*	Internal camera and flexible LED array with elastomeric sensing layer	Compliant robotic palm; evaluated through contact-area coverage

Summary

The proposed palm-scale sensor is capable of estimating the indentation depth, contact location and normal force with sub-millimetre localisation, with $R^2 > 0.999$, a MAE of 0.165 mm for the indentation depth, and an F_z MAE of 0.486 N. It has a higher force error than specialised optical tactile systems, yet its compact, modular design enables the generation of multiple tactile outputs for robotic grasping and manipulation.

8 Discussion

8.1 Key Findings and Contributions

This work presents a modular photoelastic tactile sensor for robotic palms that estimates contact force and location from a single camera image using a folded optical path. Four calibration architectures were evaluated using a labelled automated-indentation dataset. The differential-crop network achieved the lowest normal-force error, with an F_z MAE of 0.244 N. In comparison, the physics-informed two-stage pipeline achieved a 2D localisation MAE of 0.204 mm, and an indentation-depth MAE of 0.165 mm. The flattened-array MLP also achieved a competitive force MAE of 0.330 N despite removing spatial convolution, indicating that pixel intensity alone contains useful force-related information. Overall, the results demonstrate the feasibility of single-camera, transmission-based photoelastic sensing at the palm scale.

8.2 Fringe Quality and Optical Path Design

The prototyping suggested that both fringe contrast and image resolution affected the ease of use of the sensors, and that the optical path design and illumination uniformity offered significant advantages. The first iteration, based on a vertically stacked camera and lateral illumination, resulted in a low fringe contrast, even though adequate imaging resolution was achieved. The second iteration solved this problem, but not by reducing the optical path; the camera was moved to the side, and the underside of the membrane was bounced up to the lens using an angled mirror. This solved the contrast issue while also freeing up the space under the palm that a robotic hand would need for structural support and actuation. A diffused back-illumination stack (an intensity-reducing grid followed by three diffuser sheets) was also needed to eliminate discrete LED artefacts that otherwise might be misinterpreted as fringes [10, 11]. These results indicate that the main design parameters for a reliable photoelastic readout at the palm scale are illumination uniformity and the optical path geometry, rather than the sensor's resolution alone.

8.3 Design and Deployment Trade-offs

A folded optical path also represents a compromise between mechanical complexity and optical integration compactness, with a significant gain in the latter area [27, 12]. A vertically stacked design cannot accommodate actuator and transmission components under the palm, whereas placing the camera near the wrist leaves space there for these

components. The independent mounting of the illumination, polarisation and imaging components used in the second prototype enabled the material comparison described in Section 4 to be realised as candidate membranes could be changed without rebuilding the optical path [12]. It would also allow easy replacement of the camera or illumination equipment without redesigning the sensor housing, if needed in the future.

8.4 Calibration Architecture and Representation Choice

In Section 5, it is shown that across all three different representations (RGB, greyscale and flattened 1D), no particular one was dominant when compared across various tasks. The differential RGB representation achieved the best direct normal-force estimation, in line with its ability to suppress static background content compared to the raw-crop network [15, 24]. The results from the greyscale, physics-informed pipeline showed the best localisation, indicating that intensity and fringe structure are sufficient for estimating spatial contact and suggesting that colour information was not necessary for achieving high localisation accuracy in this pipeline. Even though the 1D flattened-array MLP completely dropped the spatial structure, it remained competitive in force performance, suggesting that a portion of the force-relevant signal is carried by overall pixel intensity rather than by the geometry of the fringes. Though all the architectures employed location-disjoint partitions, it is possible to compare this result with the convolutional approaches that still can be interpreted within the scope of the single experimental campaign discussed in Section 8.6 [25, 26].

8.5 Advantages

The key benefit of the sensor is that it can simultaneously acquire both stress-related and spatial contact information from a single camera image. In contrast, such information would require a scattering-based readout to achieve in a palm-sized area, as outlined in Section 3 [6, 4]. The cost of the sensor was kept at around £50 by using common, readily available materials, making it cheap to replicate and further develop. The modular design enables the sensing membrane, illumination stack, and imaging module to be replaced separately for repair and the development of future prototypes [12]. The imaging module recorded at 30 fps, which is high enough to enable the acquisition of the 25 frame-per-increment logging used during dataset collection and will allow future real-time use. Furthermore, the folded optical path allows the imaging hardware to avoid competing for space under the palm, which is crucial for potential integration into a dexterous hand [27, 5].

8.6 Limitations

Some limitations should be considered when interpreting the reported results. First, the calibration architectures can predict multiple output quantities, and their capabilities can be extended to include contact-shape classification, additional force or torque outputs, and more. Using location-disjoint partitions, the samples used for training and evaluation were drawn from the same sensor, experimental setup, and data-collection campaign. The reported performance therefore demonstrates generalisation to held-out physical contact locations within the completed experimental campaign. Still, it does not demonstrate generalisation to independently collected datasets, unseen objects, different sensor configurations, or substantially different operating conditions. In addition, no shape-labelled dataset or systematically varied lateral-force-and-torque dataset exists to evaluate such extensions, using the same spherical indenter to perform all indentations under controlled normal-loading conditions.

Second, the F_z error for the two-stage pipeline is higher than the reported errors of specialised fingertip-scale optical systems such as Insight and the stereo-tactile sensor of Kakani et al. [6, 4]. However, these systems were evaluated over different sensing areas, load ranges, and validation procedures.

Third, contact localisation using the mirror crop is subject to a transparency-related confound. Note that the optically transparent membrane not only sees deformations of the bottom surface of the membrane caused by contact, but it can also see the indenter or the object in contact with the membrane through the membrane as well. It is therefore possible to have a part of the signal for localisation that is learned directly from the object that the membrane is in contact with, rather than just from the mechanical deformation of the membrane due to contact. In the present dataset, the contact position was fully coupled to the indenter's shape and appearance, as the only object the sensor experienced was the indenter; therefore, this effect is difficult to distinguish from the others.

9 Conclusion

In this work, a modular, camera-based tactile sensor for robotic palms is presented that estimates contact force and location from photoelastic fringe patterns. The optical path is folded using an angled mirror, which separates fringe imaging from the volume beneath the sensing membrane, thereby enabling the camera to be positioned close to the wrist rather than directly under the palm. This overcomes a major packaging limitation for the wider scale-up of photoelastic tactile sensing beyond the fingertip scale. Its lightweight calibration architectures and ability to estimate force and contact position from a single camera image enabled it to estimate both simultaneously.

Comparing four different image representation calibration architectures, it was found that no single image representation was optimal for all tasks. Differential RGB inputs yielded the best results for normal force estimation. In contrast, the physics-informed pipeline yielded significantly smaller localisation errors, down to the sub-millimetre range. The findings are consistent with the hypothesis that colour information contributes to force estimation, but did not separate the effects of colour representation and architecture. Interestingly, the force prediction remained competitive with a simple multi-layer perceptron running on a flattened 1D array of pixels, without any explicit spatial or convolutional processing. This suggests that a significant amount of force-relevant information can be extracted without learning spatial relationships, which could be beneficial for deployment on resource-constrained embedded hardware. Practical benefits of the modular construction are also obtained. The sensor was built for about £50 using readily available materials, and the illumination, polariser, and imaging components could be replaced individually without reconstructing the optical path. The imaging module is currently running at 30 fps, with scope for real-time inference, after further optimisation of the calibration pipeline for embedded systems. These features make the sensor easy to replicate, maintain, and iterate on.

The findings demonstrate the ability of photoelastic fringe patterns to serve as a compact sensing modality operating at the human palm scale and carrying rich information. The LocalisationCNN achieved the best contact localisation accuracy, while the flattened-intensity representation yielded comparable results without convolutional processing and still enabled accurate force estimation. Overall, the results provide a proof of concept that the proposed modular, low-profile optical arrangement can be used in conjunction with learning-based calibration for photoelastic tactile sensing at the palm scale under the experimental conditions explored in this study.

10 Future Work

The limitations identified in Sec.8.6 suggest a few directions for future work, as detailed below.

The most direct extension is to the dataset, not the architecture, since the calibration pipelines are already designed to produce multiple joint outputs and do not need to be restricted to normal-force, single-shape contacts. A variety of indenter shapes – flat, edged, multi-point and the spherical indenter employed here – will allow contact-shape labels to be collected for the first time. This would enable the use of the above-mentioned multi-task heads with shape output, along with force and location outputs, since this has already been implemented for photoelastic sensors at the finger scale [10]. The addition of controlled lateral loading and torque to the rig would yield systematically varied multi-axis loading force data for adequate characterisation and validation of three-dimensional force sensing, which is not possible using the accidental F_x and F_y data collected during the normal indentation process [6, 28].

The direction of future work is also given by the identified transparency-related confound in Sec. 8.6. The optically transparent sensing membrane now allows the mirror crop to view the object being contacted directly through the membrane and to observe the actual contact-induced deformation of the underside of the membrane. A controlled ablation study could compare localisation performance across dataset variants in which the visible indenter is digitally or physically occluded. Alternatively, the membrane could be redesigned to incorporate a thin, controlled opacifying layer so that fringes remain visible while preventing a direct optical path through the membrane to the mirror. Before the present results can be extrapolated to previously unseen objects, which will not be as visually unique as the single indenter used for training, it is important to determine how much of the reported localisation accuracy is due to this direct visibility, as opposed to the mechanical deformation which the sensor is designed to measure.

Measurement of the membrane's durability after repeated indentations and characterisation of the loss of optical clarity over time following surface damage could provide practical guidelines for the maintenance and lifetime of the membrane, particularly in modular architecture, where individual membranes are to be replaced. An interesting direction would be to study the sensor's behaviour when subjected to a sustained load at a single location to see if the viscoelastic hysteresis seen in the five repeats per grid point further accumulates over a longer period of time [10].

Finally, it is important to demonstrate that the sensor is useful in real-world grasping or manipulation scenarios before the results presented herein can be considered representative of deployment scenarios. The type of testing would involve varying lighting, different or new contact geometries, multi-region or simultaneous loading, and uncontrolled contact orientation, none of which are present in the structured indentation used here. It would also allow the first test of the generalisation of the calibration pipelines to all varieties of objects a robotic palm might encounter in operation.

References

[1] 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, 2017.

[2] Y. Serrien, "Real-time multi-task tactile sensing using photoelastic fringe patterns and deep learning," Master's thesis, Delft University of Technology, Delft, The Netherlands, 2025. M.Sc. thesis.

[3] A. Alspach, K. Hashimoto, N. Kuppuswamy, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," *arXiv preprint arXiv:1904.02252*, 2019.

[4] V. Kakani, X. Cui, M. Ma, and H. Kim, "Vision-based tactile sensor mechanism for the estimation of contact position and force distribution using deep learning," *Sensors*, vol. 21, no. 5, p. 1920, 2021.

[5] S. Q. Liu and E. H. Adelson, "A passively bendable, compliant tactile palm with RObotic modular endoskeleton optical (ROMEO) fingers," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 691–697, IEEE, 2024.

[6] H. Sun, K. J. Kuchenbecker, and G. Martius, "A soft thumb-sized vision-based sensor with accurate all-round force perception," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 135–145, 2022.

[7] J. Hu, S. Cui, S. Wang, C. Zhang, R. Wang, L. Chen, and Y. Li, "GelStereo Palm: A novel curved visuotactile sensor for 3-D geometry sensing," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 11, pp. 10853–10863, 2023.

[8] J. Hu, S. Cui, S. Wang, R. Wang, and Y. Wang, "Active shape reconstruction using a novel visuotactile palm sensor," *Biomimetic Intelligence and Robotics*, vol. 4, p. 100167, 2024.

[9] V. N. Dubey and R. M. Crowder, "A dynamic tactile sensor on photoelastic effect," *Sensors and Actuators A: Physical*, vol. 128, no. 2, pp. 217–224, 2006.

[10] D. Mukashev, N. Zhuzbay, A. Koshkinbayeva, B. Orazbayev, and Z. Kappassov, "PhotoElasticFinger: Robot tactile fingertip based on photoelastic effect," *Sensors*, vol. 22, no. 18, p. 6807, 2022.

[11] M. Mitsuzuka, J. Takarada, I. Kawahara, R. Morimoto, Z. Wang, S. Kawamura, and Y. Tajitsu, "Application of high-photoelasticity polyurethane to tactile sensor for robot hands," *Polymers*, vol. 14, no. 23, p. 5057, 2022.

[12] M. Lambeta, P.-W. 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.

[13] Q. Cong, S. Oh, W. Fan, S. Luo, K. Althoefer, and D. Zhang, "Taceva: A performance evaluation framework for vision-based tactile sensors," *Advanced Intelligent Systems*, vol. 8, no. 4, p. e202501179, 2026.

[14] I. Andrussow, H. Sun, K. J. Kuchenbecker, and G. Martius, "Minsight: A fingertip-sized vision-based tactile sensor for robotic manipulation," *Advanced Intelligent Systems*, vol. 5, no. 8, p. 2300042, 2023.

[15] M. R. I. Prince, S. Athar, P. Zhou, and Y. She, "Tacscope: A miniaturized vision-based tactile sensor for surgical applications," *Advanced Robotics Research*, vol. n/a, no. n/a, p. e202500117, 2025.

[16] Q. Zhang, Z. Zuo, H. Wang, B. Liu, Y. Yilihamu, and L. Wen, "Utact: Underwater vision-based tactile sensor with geometry reconstruction and contact force estimation," *Advanced Robotics Research*, vol. 2, no. 3, p. e202500091, 2026.

[17] J. Ren, W. Du, Y. Dong, N. Zhang, H. Guo, B. Shi, J. Zou, and G. Gu, "Gellight: Illumination design, modeling, and optimization for camera-based tactile sensor," *Advanced Intelligent Systems*, vol. 7, no. 5, p. 2400596, 2025.

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[19] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proceedings of the 30th International Conference on Machine Learning*, pp. 1310–1318, 2013.

[20] L. Massari, G. Fransvea, J. D'Abbraccio, M. Filosa, G. Terruso, A. Aliperta, G. D'Alesio, M. Zaltieri, E. Schena, E. Palermo, *et al.*, "Functional mimicry of ruffini receptors with fibre bragg gratings and deep neural networks enables a bio-inspired large-area tactile-sensitive skin," *Nature Machine Intelligence*, vol. 4, no. 5, pp. 425–435, 2022.

[21] L. Massari, E. Schena, C. Massaroni, P. Saccomandi, A. Menciassi, E. Sinibaldi, and C. M. Oddo, "A machine-learning-based approach to solve both contact location and force in soft material tactile sensors," *Soft robotics*, vol. 7, no. 4, pp. 409–420, 2020.

[22] Z. Chen, N. Ou, X. Zhang, Z. Wu, Y. Zhao, Y. Wang, E. S. Papastavridis, N. Lepora, L. Jamone, J. Deng, *et al.*, "Training tactile sensors to learn force sensing from each other," *Nature Communications*, vol. 17, no. 1, p. 2101, 2026.

[23] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation," *Peerj computer science*, vol. 7, p. e623, 2021.

[24] W. Fan, H. Li, Y. Xing, and D. Zhang, "Design and evaluation of a rapid monolithic manufacturing technique for a novel vision-based tactile sensor: C-Sight," *Sensors*, vol. 24, no. 14, p. 4603, 2024.

[25] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," *Patterns*, vol. 4, no. 9, 2023.

[26] L. Sasse, E. Nicolaisen-Sobesky, J. Dukart, S. B. Eickhoff, M. Götz, S. Hamdan, V. Komeyer, A. Kulkarni, J. M. Lahnakoski, B. C. Love, *et al.*, "Overview of leakage scenarios in supervised machine learning," *Journal of Big Data*, vol. 12, no. 1, p. 135, 2025.

[27] J. Zhao and E. H. Adelson, "GelSight Svelte: A human finger-shaped single-camera tactile robot finger with large sensing coverage and proprioceptive sensing," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1646–1653, IEEE, 2023.

[28] W. Li, A. Alomainy, I. Vitanov, Y. Noh, P. Qi, and K. Althoefer, "F-TOUCH sensor: Concurrent geometry perception and multi-axis force measurement," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4300–4309, 2021.

Acknowledgements

The path to completing this dissertation has not been straightforward, presenting both professional challenges and personal difficulties. I would like to thank my supervisor, Prof Kaspar Althoefer, whose critiques and ideas shaped the direction of this project from the outset.

I am equally grateful to Dr Abu Bakar Dawood and Dr Dalia Osman for their time and effort, and for the unfailing support and constructive feedback that did so much to improve the quality of this work.

My thanks also go to the wider staff of the Centre for Advanced Robotics @ Queen Mary (ARQ), of which my supervisors are themselves members, whose resources and assistance have been invaluable throughout.

A Appendix

A.1 Code Repository

The source code and supporting scripts developed for the machine-learning methods and quantitative analysis presented in this thesis are available in the following public GitHub repository:

https://github.com/Spades2002/Photoelastic-Palm-Sensor

The repository contains the implementations of the four modelling approaches investigated in this work: the raw-crop dual-ResNet architecture, the differential-crop force-conditioned ResNet architecture, the differential intensity-based MLP, and the physics-informed two-stage CNN architecture. Each method is organised in a separate directory containing the corresponding model implementation, preprocessing pipeline, training procedure, evaluation workflow, and supporting documentation. The repository also includes a dedicated `Graphs` directory containing the Python scripts used to generate the quantitative figures presented in this thesis from the recorded model predictions. A repository-level README provides an overview of the complete codebase, the organisation of the individual modelling approaches, and the associated analysis scripts.