

High-Sensitivity Vision-Based Tactile Sensing Enhanced by Microstructures and Lightweight CNN

Mayue Shi^{1*}, Yongqi Zhang¹, Xiaotong Guo¹ and Eric M. Yeatman¹

¹Department of Electrical and Electronic Engineering, Imperial College London, London, UK.

*Email: m.shi16@imperial.ac.uk.

Abstract:

Tactile sensing is critical in advanced interactive systems by emulating the human sense of touch to detect stimuli. Vision-based tactile sensors (VBTSs) are promising for their ability to provide rich information, robustness, adaptability, low cost, and multimodal capabilities. However, current technologies still have limitations in sensitivity, spatial resolution, and the high computational demands of deep learning-based image processing. This paper presents a comprehensive approach combining a novel sensor structure with micromachined structures and an efficient image processing method, and demonstrates that carefully engineered microstructures within the sensor hardware can significantly enhance sensitivity while reducing computational load. Unlike traditional designs with tracking markers, our sensor incorporates an interface surface with micromachined trenches, as an example of microstructures, which modulate light transmission and amplify the variation in response to applied force. By capturing variations in brightness, wire width, and cross pattern locations with a camera, the sensor accurately infers the contact location, the magnitude of displacement and applied force with a lightweight convolutional neural network (CNN). Theoretical and experimental results demonstrated that the microstructures significantly enhance sensitivity by amplifying the visual effects of shape distortion. The sensor system effectively detected forces below 10 mN, and achieved a millimetre-level single-point spatial resolution. Using a model with only one convolutional layer, a mean absolute error (MAE) below 0.05 mm have been achieved. Its soft sensor body ensures compatibility with soft robots and wearable electronics, while its immunity to electrical crosstalk and interference guarantees reliability in complex human-machine environments.

Keywords:

Vision-based tactile sensor (VBTS), microstructure, convolutional neural network, stretchable, MEMS.

1. Introduction

Tactile sensing is a critical aspect of intelligent interactive systems [1]. It mimics the human sense of touch, enabling the detection of stimuli such as pressure and vibration [2]. Many applications have been developed in various fields [3], such as robotics [4], healthcare monitoring [5], [6], and prosthetics [7], [8]. However, current tactile sensing technologies encounter significant challenges, especially in sensitivity, resolution and durability. Additionally, for large-scale sensing, which is critical for robots and prosthetics applications, current sophisticated sensing systems based on sensor arrays often have to face dramatical increase of complexity, electrical crosstalk and interference, and usually unacceptable cost. It is necessary to develop innovative solutions to enhance tactile sensing capabilities.

Vision-based tactile sensors (VBTSs), which utilise visual information to infer tactile properties, are promising to address the limitations in current tactile sensing. By utilising visual data, vision-based tactile sensors can effectively detect location and magnitude of pressure, characterise surface textures, and operate in a wide range of scenarios. This innovative method has been investigated for robotic perception [1], offering rich information, high robustness and adaptability, low cost, and multimodal sensing capability [9]. Moreover, VBTSs are compatible with many computer vision methods, such as convolutional neural networks [2]. Recently, the rapid development of artificial intelligence technology has effectively boosted the development of tactile sensors [10]. Computer vision methods can be used to extract features from the images captured by the camera module of the sensor. These features can then be used to estimate contact position and force distribution.

The structure of VBTS often contains a sensor body and camera, among which the design of sensor body is critical for the improvement of perception performance, while the camera module remains relatively standardised. The optimisation design of the sensor body generally involves several aspects, including novel structures, advanced materials and fabrication processes. With the rapid improvement of hardware designs, the last decade has witnessed the development of the VBTS towards miniaturization, high performance and multimodality [11]. Current VBTSs are mainly based on two principles: displacement of markers with deformation of the sensor [12], [13], [14], [15], [16], and irregular refraction and reflection of light induced by surface deformation [17], [18], [19]. For example, GelForce was a marker-based sensor designed to measure the surface traction field on a robotic hand [15], [16]. It contained two layers of spherical markers embedded within a transparent silicone rubber sensor body. In addition, the TacTip family realised superresolution tactile sensing for localization tasks with a 3D printed sensor body with markers, utilising pins on the sensor body to mimic the function of intermediate ridges within the human fingertip [13]. Early GelSight sensors captured high-

resolution geometry without the uses of markers. They utilised photometric stereo to reconstruct the depth map and to infer local force [19], [20]. Later versions of GelSight introduced markers in the reflective membrane to enhance the capability to measure force and torque loads [18], [21]. In addition to advances in sensor body design, improvements in camera systems and advanced image processing methods further improved the performance of VBTSSs. A multi-camera design was presented by Trueeb et. al [22] to overcome the limited field of view inherent in single-camera systems. By monitoring the particle patterns in the sensor body, contact force distribution was measured with a machine learning method. A transfer learning algorithm was also explored to reconstruct normal force distribution and facilitate knowledge transfer across different sensor gels. [23]. Evetac is a novel event-based optical tactile sensor designed to enhance temporal resolution and detect vibrations up to 498 Hz, leading to more adaptive and precise robotic manipulation. It processed measurements online at 1000 Hz, using efficient algorithms to track deformation and reduce data rates compared to traditional RGB sensors [24]. Recently, large-scale tactile sensing methods for robotic manipulators have gained attention. For instance, a study demonstrated the use of electroluminescent panels combined with a deformable soft skin to localize forces and estimate their magnitude in multipoint contact scenarios, achieving millimetre-level localisation accuracy [25].

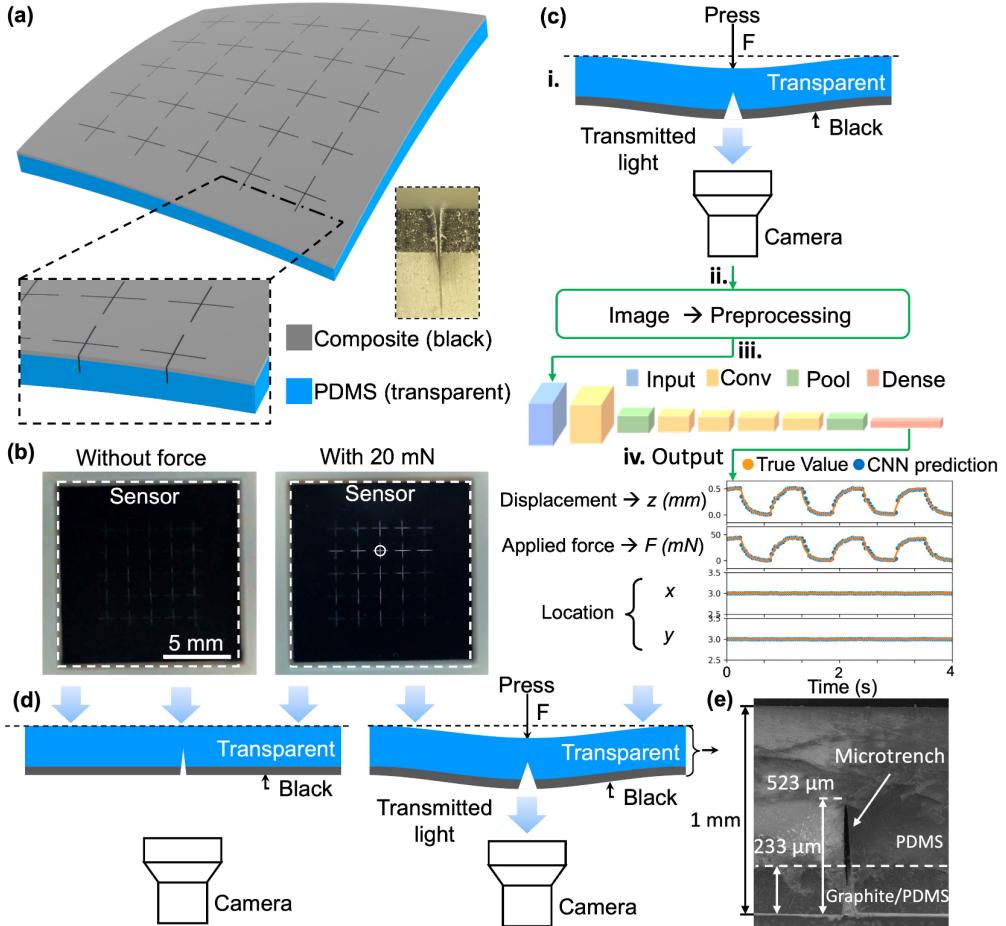


Figure 1: Structural schematic diagram and system design. (a) Structural design of the contact surface, and a sectional view showing trench structures in the film. (b) Experimental result with an applied force of 20 mN at the cross centre in row 2, column 3. Diameter of contact area: 1 mm. (c) System design. i. Contact surface with micro trench patterns allows light to transmit under pressure. ii. Camera captures images with transmitted light patterns and then iii. the images are pre-processed, and iv. feed into a light-weight CNN for analysis. v. Three outputs of CNN represent vertical displacement, and the location of the contact point on the sensor surface. (d) Principle of structure-enhanced vision-based sensor. Left: without pressure; right: with pressure. Deformation of surface under pressure allows light to transmit from open micro trenches, forming specific patterns captured by the camera. (e) Characterisation of a micro trench with SEM microscope.

Despite some studies exploring innovative structural designs for VBTSSs, the potential performance benefits of integrating microfabrication approaches and micromachined features into these sensors have not been investigated, although they have been well established for both soft and rigid materials, and numerous micro-electromechanical systems (MEMS). Recently, we have successfully developed the fabrication process of a microstructure-enhanced sensor body [26], and initially demonstrated its potential to achieve high precision in sensing of both location and applied force. In the current study, we comprehensively developed this system from hardware to data processing, to infer the displacement, location and magnitude of applied force with customised computer vision methods. Specifically, the ultra-light model based on CNN was developed for inferring sensor outputs from images. The performance of the VBTSS was also evaluated, which demonstrated that the model can effectively realise multimodal sensing (**Figure 1**). With micromachined cross-shaped structures on the stretchable sensor body and the computer vision method developed in this paper, the properties of the applied force including magnitude and contact location can be measured with both high precision and sensitivity.

2. Results

2.1. Structural design and working principle

As shown in **Figure 1a**, the stretchable sensor body was composed of a transparent polydimethylsiloxane (PDMS) layer and a thin black composite layer. A grid of 5×5 cross-shaped trench structures was precisely fabricated using ultraviolet laser cutting, with a depth of around halfway through the sensor body film (**Figure 1e**). The distance between adjacent cross centres is 2 mm, with each cross measuring 1.5 mm in both length and width. The sensor body is clamped using a 3D-printed frame with a 1.6x1.6 cm square window. The architecture of the

sensor system, as depicted in **Figure 1c**, involves image acquisition and preprocessing, followed by lightweight CNN processing to learn the patterns and features necessary for prediction of tactile measures, including displacement in depth (z -coordinate), applied force and location coordinates (x, y). As shown in the figure, the comparison between true values and CNN predictions demonstrates the system's high accuracy.

Figure 1d illustrates the working principle of this sensor that leverages the deformation of microstructures to measure applied force. When applying pressure on the transparent side, the deformation of the film leads to an increase of micro trench width and luminous flux, and images are captured by a camera under the sensor body. For instance, the images of our sensor body clearly showed the distinguishable features under an applied force of only 20 mN (**Figure 1b**), which indicates its high sensitivity. With the diaphragm structure and elastic sensor body, the sensor recovers spontaneously after deformation. We have modelled the deformation with finite element analysis, which indicates that the micro trenches effectively enhance film deformation under the same force.

2.2. Theoretical modelling and simulation

We developed a comprehensive mathematical model to analyse the bending behaviour of a thin PDMS film with fixed boundaries and uniformly distributed V-shaped notches (**Figure 2**). With this model, we derived the extension of trench width (Δd) under an applied force, which is critical for modulating light transmission and amplifying the variation. This model simplifies the complex behaviour of the sensor body film by representing it as an array of fixed-fixed beams. Each beam bends independently in response to an applied load, assuming that cross-sections remain undeformed during bending. A detailed modelling process is presented in the **Supplementary Note 1**.

As an example, for location A, the deformations at the left and right edges of the notch (**Figure 2b**) can be expressed as:

$$\Delta d_A = \Delta d_{A_1} + \Delta d_{A_2} = \frac{49M_A}{22wE\alpha} + \frac{17M_A}{22wE\alpha} \quad (3)$$

Here, note that E denotes the effective Young's Modulus of the film, determined through experimental calibration. This method can, similarly, be applied to other intermediate locations, using their respective moment diagrams to guide the geometric representation of the stress-free areas.

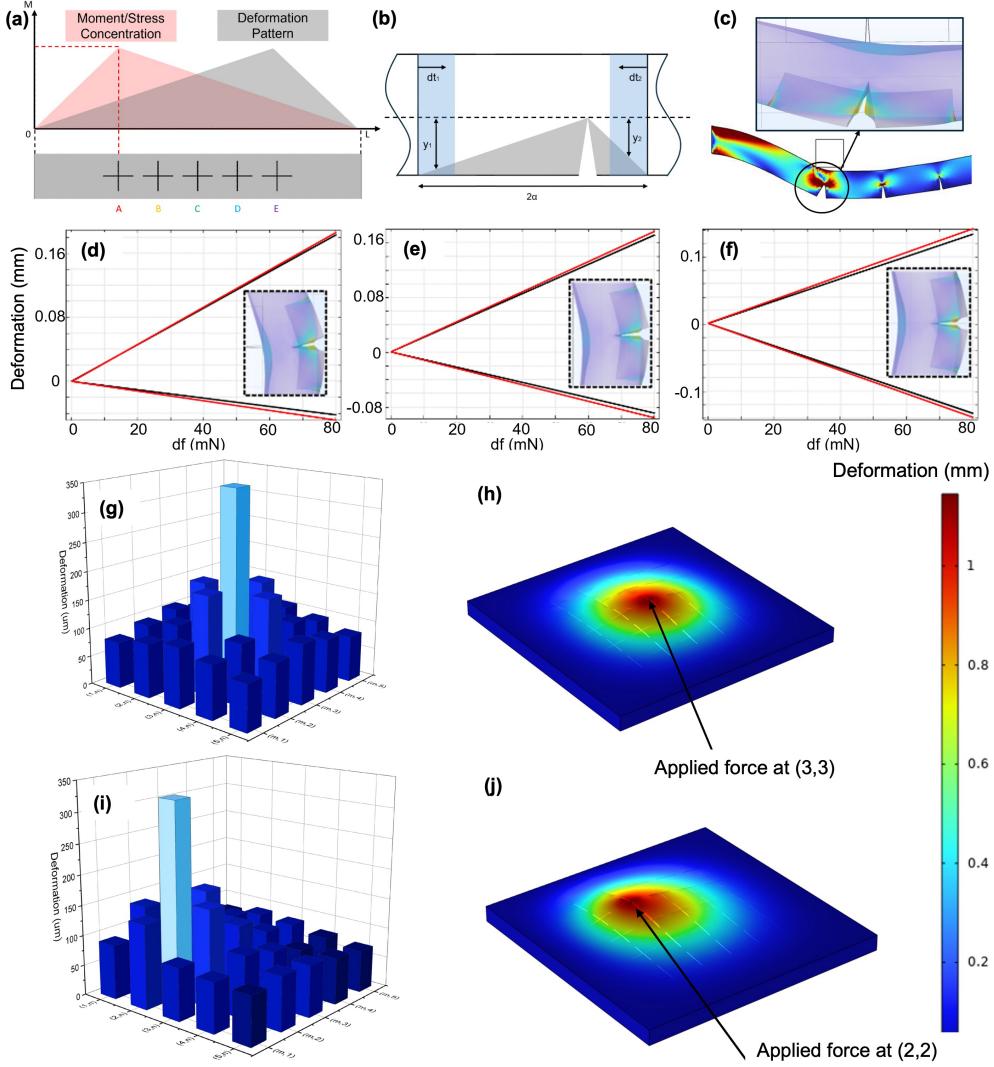


Figure 2. Modelling and simulation analysis. (a) to (d) representation of the stress-free area using intermediate location A as an example. Unlike the central location, the stress-free area is not an isosceles triangle. Instead, it is shaped in accordance with the corresponding moment diagram. The stress-free area's geometry reflects the moment diagram, showcasing a shape that accurately captures the stress-free region for this location. (e), (f), and (g) present the deformation pattern of the notched beam at locations C, B, and A simulated using COMSOL Multiphysics. The load range analysed was from 0 to 80 mN. Comparison between the simulated and modeled (red) deformations demonstrates our model consistently predicted a slightly larger deformation at both edges, with discrepancies within 13% across varying applied loads. (g)-(j) Deformation of all 25 micro trenches under an applied force of 60 mN at location (3, 3) and location (2, 2).

To validate the mathematical model, we compared the FEA results with the analytical predictions (**Figure 2e,f,g**). Using COMSOL Multiphysics, we performed simulations over a range of applied loads (0 to 80 mN), observing deformation patterns that closely matched our

theoretical predictions. Further FEA was conducted (**Figure S3**) to model the deformation and stress distribution of the sensor body under a 60 mN force applied at the central location. The analysis revealed significant deformation around the contact point, affecting the micro trench's luminous flux. To illustrate the deformation patterns across the 5×5 array, simulations were conducted for two scenarios (**Figure 2g to j**): a 60 mN force applied at the central location (3, 3) and at an off-center location (2, 2). As expected, demonstrated by Figure 6, deformation patterns were radially symmetrical for the central location and varied for the off-centre location, with further locations undergoing less deformation.

2.3. Experimental set-up and data collection

The sensor body was prepared with two materials: transparent PDMS elastomer and black graphite/PDMS composite. Detailed fabrication procedures the sensor body are provided in **Figure S5 and Methods**. According to SEM measurement (**Figure 1e**), the total thickness of the sensor body film is 1mm, while the composite layer is $\sim 230 \mu\text{m}$. The depth of the micro trench was $\sim 523 \mu\text{m}$.

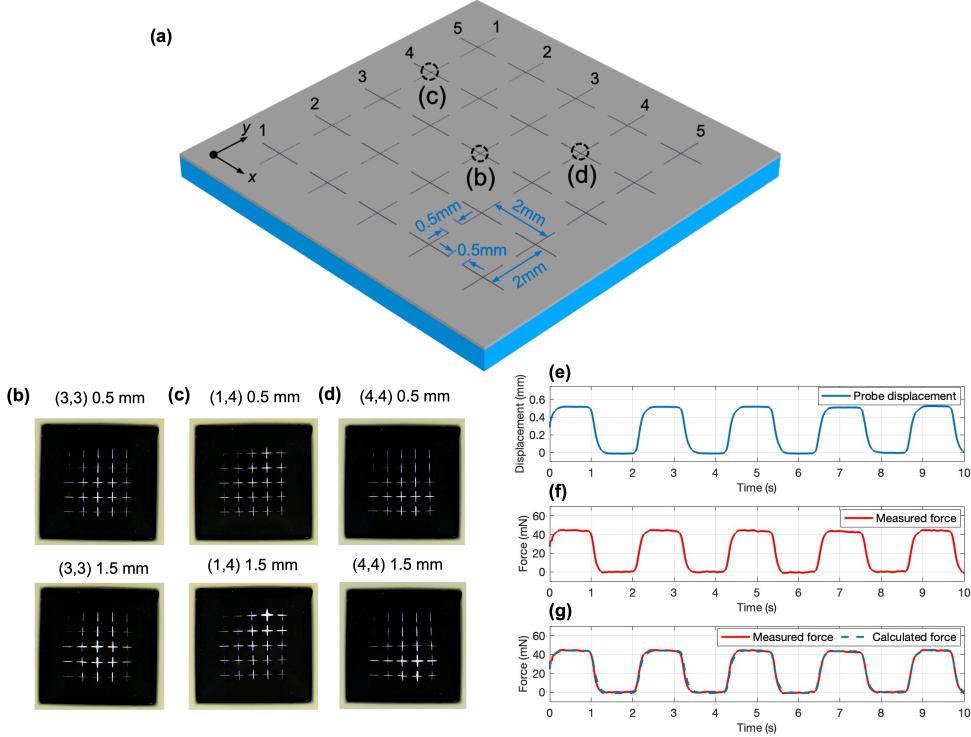


Figure 3: Sensor body and image captured under various conditions. (a) Structure and parameters of the sensor body. (b)(c)(d) Optical images captured using camera. The data above each image indicates the contact location and the displacement of the linear motor toward the sensor. For example, (3, 3) indicate the applied force was at the cross centre in row 3, column 3. Diameter of contact area: 1 mm. (e)(f)(g) represent the force - displacement calibration of a testing point (3,3). (e) Displacement and (f) force was measured to calibrate the coefficient k_{33} .

(g) Comparison of the measured force and the calculated force according to the displacement over time ($k_{33}=85.4$ mN/mm).

The performance of the sensor was tested with a bespoke experimental setup, as shown in **Figure S6** and **Methods**. During image data collection, a commercial Logitech C922 camera (1920×1080 pixels resolution) was used to measure brightness changes at each sensing point during force applications. Displacements of 0–0.5 mm, 0–1.0 mm, and 0–1.5 mm were applied against the sensor surface, with all sensing points tested repeatedly using the same characterization process. The VBTS was tested under environmental light. We first tested the sensor with a single contact point located at the centre of each cross, after which the points between the two centres were tested. Images captured under these conditions showed clear distinctions. The results from three typical testing points of (3,3), (4,3) and (3.5,4.5) are shown in **Figure 3**. The images indicate that crosses close to the contact point have high luminous flux and wide trenches. Additionally, the deformation of the film led to the displacement of crosses in images. This result was predicted by finite element simulation as well.

A simple quantitative image analysis is shown in **Figure S4**. We calculated the average brightness of the square central region (60×60 pixels) of the images in Figure 8b, to reflect the response of the sensor to different vertical displacements. The calculation is performed using value channels of images in the HSV (hue, saturation, value) colorspace, in which the value describes the brightness. The figure clearly indicated that the average brightness increased with the displacement. Specifically, when the displacement varied from 0.5 mm to 1.5 mm, the average brightness increased to 3.2 times that of 0.5 mm displacement, which suggests the outstanding sensitivity of our sensor in this range, even when evaluated solely through the metric of average brightness. If we only focus on the cross-trench region, the variation of brightness was even higher. This is especially beneficial for further in-depth image processing based on machine learning models to achieve higher sensitivity.

The applied force can be calculated based on vertical displacement (z -direction). According to theoretical analysis and simulation [26], the relationship can be approximately expressed as $F(z) = k_{xy} \times z$, where the coefficient k_{xy} represents the proportionality between vertical displacement (z) and the applied force (F) at location (x,y) . To calibrate the coefficient k_{xy} , we measured the applied force with a high-precision scale, as described in **Methods**. As an example, the measurement results from the point (3,3) is shown in **Figure 3e** to g. According to the measured force and probe displacement, a linear curve fit was performed, and the resulting k_{33} value was calibrated as 85.4 mN/mm. The coefficient of determination R^2 was calculated as 0.9944 at this point, which validated the linear model is a good approximation of force - displacement relationship.

2.4. CNN model architecture and evaluation

We have developed an ultra-light CNN model architecture to infer the location and magnitude of applied force, as shown in **Figure 4**. We tested models comprising 1, 3, 5 and 7 convolutional layers. As an example, the model with 5 convolutional layers (**Figure 4b**) has a total of 406,659 parameters. It begins with preprocessing layers that rescale the pixel values. The image matrix values were rescaled to a range of 0 to 1 by dividing by 255, standardising the inputs and facilitating faster convergence. The convolutional layers progressively extract features, each followed by batch normalisation and ReLU activation to enhance learning efficiency and stability. The first convolutional layer outputs a shape of (54, 54, 64), while subsequent layers further refine the features, culminating in a (10, 10, 128) output after the fifth layer. The architecture also includes max pooling and average pooling layers to reduce spatial dimensions and computational complexity. Finally, the output is flattened and passed through a dense layer to predict the x, y, and z coordinates of the applied force.

The optical information collected by the camera was pre-processed before being passed to the CNN model (**Figure 1c**). Specifically, the collected dataset included 225,000 frame images captured by the camera, each captured with applied forces at various locations and magnitudes. Each image was cropped into 25 small images of 60×60 pixels, then was input as 25 channels (**Figure 4a**). To further enhance the diversity of the training data, data augmentation techniques, including random translation and rotation, were employed. This approach helped to mitigate overfitting and improve the model's robustness and generalisation capabilities. The training process was conducted on a high-performance computing cluster, allowing for efficient handling of the large dataset. This process used 70% of the images as training set and 15% of the images as validation set, while the resting 15% was the test set.

Figure 4c demonstrates the prediction results of 100 sample images in the test set with our 5-convolutional-layer model, which reflects that the CNN model performed well in predicting the applied forces' x, y, and z coordinates. The predictions for the x and y coordinates are particularly accurate. The z coordinate predictions are slightly less precise but still follow the overall trend of the true values closely. These results suggest that the CNN model is effective for predicting the location and magnitude of applied forces, with very high accuracy for the x and y coordinates and reasonable accuracy for the z coordinate.

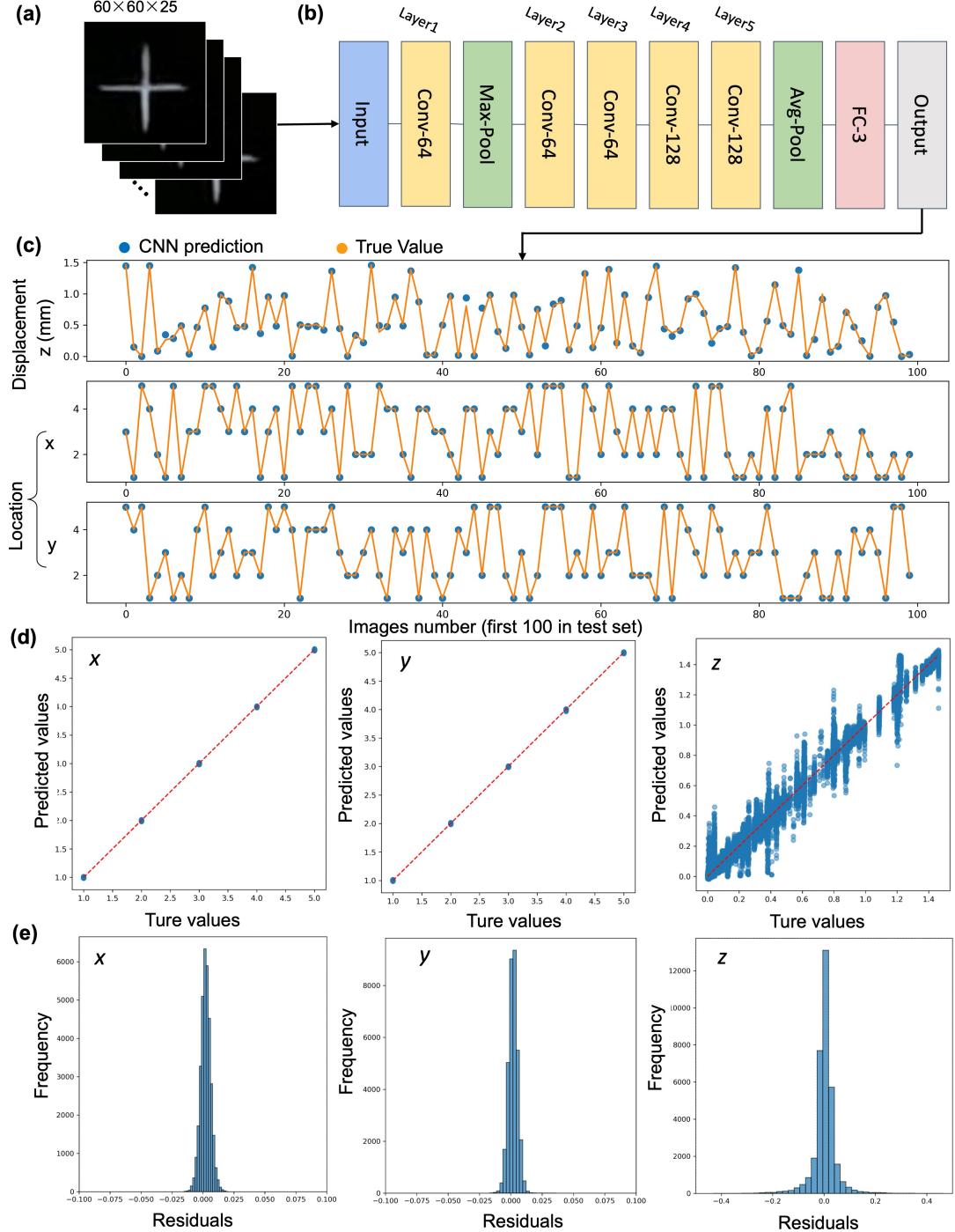


Figure 4: Architecture of the CNN model. (a) pre-processed image as input. (b) Model architecture with 5 CNN layers. (c) A comparison between the predicted values by the CNN model and the true values for the first 100 images in test set.

To evaluate the model's performance, we conducted a series of tests using the test set, containing 15% of the images of the entire dataset. The results are shown in **Figure 4d** and e.

Figure 4d presents scatter plots comparing the predicted values to the true values. Each subplot

shows a clear correlation between the predicted and true values, indicated by the close alignment of the points along the red dashed line, which represents perfect prediction (i.e., predicted values equal to true values). Consistent with the initial analysis of the results in **Figure 4c**, the scatter plot for the x - and y - coordinates (left subplot) demonstrates a nearly perfect linear relationship, with most points falling on the line, indicating high accuracy in predicting the location of the applied force. For the z -direction or displacement (right subplot), there is more spread around the line, indicating higher variance in the predictions. However, the overall trend still shows a strong correlation between the predicted and true values.



Figure 5: Error distribution for location resolution across different positions of the sensor.

Figure 4e displays histograms of the residuals, which reflect the differences between the predicted and true values. The residuals for the x and y coordinates (left and middle subplots) are narrowly distributed around zero, with the majority of residuals being very close to zero. This further confirms the high accuracy of the model for these coordinates. The residuals for the z coordinate (right subplot) have a wider distribution, suggesting that there is more error in

the predictions for the z-coordinate compared to the x and y coordinates. Nevertheless, 95% of the prediction errors fall within a narrow range. The peak of the distribution is still centred around zero, indicating that the errors are unbiased.

Figure 5 displays error distribution plots for our sensor, showing the *x-y* error patterns at 25 positions. In each subplot, the blue dots indicated individual error measurements, and the error ellipses represented the 95% confidence interval of the error. This figure shows how error varies across different positions on the sensor, with most distributions concentrated around the origin, indicating relatively low errors. However, the spread of points in certain positions shows greater variability in error, suggesting localized differences in tactile sensing accuracy across the sensor.

In general, this model containing 5 CNN layers demonstrates excellent performance in predicting the location and magnitude of applied forces, particularly for the x and y coordinates. The slightly higher error in the z-coordinate predictions suggests that there may be room for further optimization of the model or the training process to improve accuracy in this dimension. Despite this, the model's performance is robust and effective for precise tactile sensing.

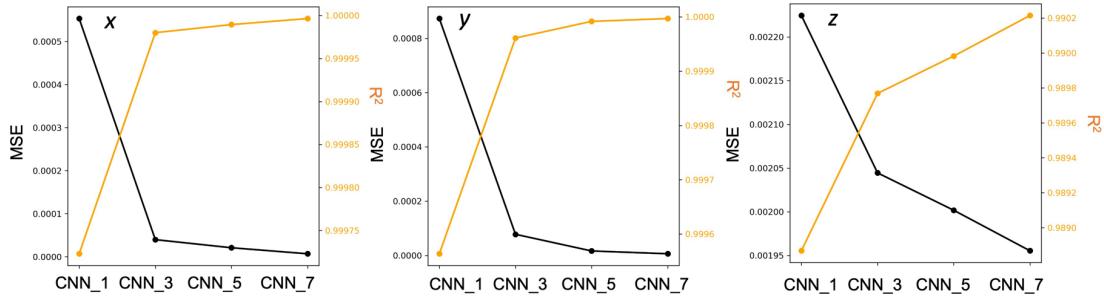


Figure 6: Comparison of performance metrics for different CNN models with varying numbers of convolutional layers (1, 3, 5, 7 layers).

We tested models with different numbers of convolutional layers from 1 to 7 layers.

Figure 6 presents the Mean Squared Error (MSE) and R² values for different CNN models with varying numbers of layers (1, 3, 5, 7 layers). **Table S1** shows a complete comparison of performance metrics for different models. The results indicates that increasing the number of CNN layers generally improves the model's performance, as indicated by decreasing MSE and increasing R² values. While performance improves up to 5 layers for the x coordinate and continues to improve up to 7 layers for the y and z coordinates, there may be diminishing returns, particularly for the x and y coordinate.

Table 1 Comparison with other works

Reference	Optical System	Image Resolution	Processing Model	Localization Error	Force/Depth Error
Duong[12]	2 fish-eye cameras	640×480	Finite element model	Mesh 18×9.5mm	>7mm depth error

Fu [25]	1 wide-angle camera	1280×720	Statistical fitting	6.22-7.19 mm	9.3%-11.1%
Gomes[17]	LED, M12 wide-lens	/	Projection model	Average 5mm	/
This work	1 web camera	1280×720	Lightweight CNN	MAE<0.04 mm, mesh 2×2 mm	MAE<0.03mm

Overall, our model achieved high precision. Due to the introduction of microstructures, the light flux variations caused by surface deformations on the sensor were amplified effectively. This amplification makes changes in graphical features more pronounced, allowing even shallow neural networks to effectively capture these features. Consequently, this reduces the computational and storage resource requirements, enabling efficient feature extraction with lightweight CNN architectures. This dual advantage highlights the efficacy of our design in balancing performance with computational efficiency. **Table 1** provides a comparative analysis of our results against other VBTS studies. The table highlights differences in optical systems, image resolutions, processing models, and errors in localization and force/depth measurements.

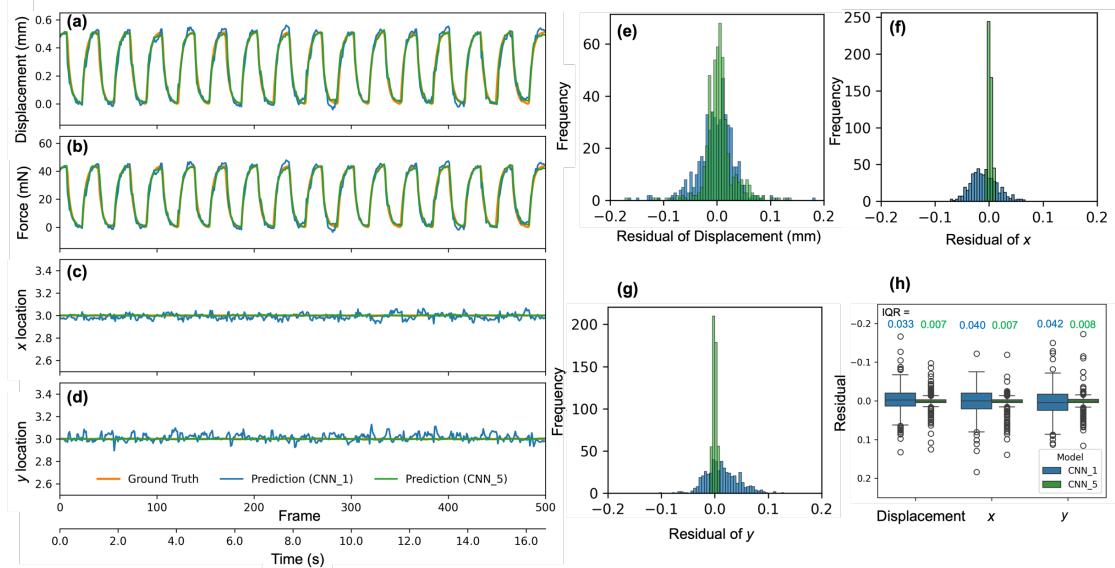


Figure 7: Experimental testing results and residual analysis at location (3,3). (a) Displacement (z -coordinate), (b) calculated force, (c) x and (d) y -coordinate results. (e) – (h) shows the residual distributions and boxplots to evaluate prediction accuracy for this sensor system.

To validate the accuracy of the VBTS system, continuous testing was performed at point (3, 3). **Figure 7** shows the results of 500 consecutive frames captured at a rate of 30 frames per second. The predicted displacement and force closely match the ground truth, as indicated by the overlapping curves. To assess the performance of models with different depth, two models were evaluated: one with a single CNN layer (CNN_1) and another with five CNN layers (CNN_5). The applied force was calculated with $k_{33}=85.4$ mN/mm, as discussed previously.

The predicted displacement and force from both CNN_1 and CNN_5 aligned closely with the ground truth. This strong correlation validated the system's high temporal accuracy and reliability during continuous testing. Both CNN_1 and CNN_5 perform well, but CNN_5 exhibits slightly better alignment with the ground truth, while the CNN_1 already provide a high accuracy, especially for x and y locations. These results were consistent with the comprehensive evaluation in **Table S1**. **Figure 7h** shows the boxplot for residuals for displacement, x , and y locations show that CNN_5 consistently performs better than CNN_1, as indicated by its lower interquartile ranges (IQRs).

3. Conclusions

In this paper, we developed an advanced VBTS enhanced by novel microstructures and light-weight CNN neural networks. This innovative sensor design, featuring a black composite layer with 25 cross trenches, modulates the light luminous flux based on the magnitude and location of the applied force. The sensor images, captured by a camera, provide the input for our ultra-light CNN model architecture to accurately infer the location, displacement and applied force. We evaluated the models consists of 1 to 7 convolutional layers. Based on our analysis, the convolutional layers effectively extract relevant features from the input images, facilitating precise predictions. The model's structure balances computational efficiency with high accuracy, making it suitable for real-time applications in various IoT and edge-computing fields. Our VBTS demonstrated remarkable performance, providing a millinewton-level sensitivity and a millimeter-level resolution. The high performance, combined with the sensor's soft sensor body, makes it promising for seamless integration into advanced soft robotics and wearable electronics. The sensor's compatibility with computer vision methods enhances its potential for integration with and further enhanced by advanced AI systems.

In future research, our sensor system can be expanded for broader applications in robotic sensing and human-machine interfaces. These cross-shaped micro trenches demonstrated here are one example of use of microstructures for enhancing vision-based tactile sensing. Other microstructures, such as more complex three-dimensional geometries which can be realised through MEMS fabrication processes, hold significant potential for further improving the sensor's sensitivity and resolution. Additionally, its potential to perform multi-point tactile sensing will be explored. Our results from both theoretical and experimental perspectives pave the way for the VBTS to play a significant role in advancing soft robotics and other human-machine interfaces requiring precise tactile sensing. We have demonstrated that the combination of innovative sensor design, advanced fabrication and powerful machine learning algorithms is particularly promising for bringing significant advancements in tactile sensing technologies.

4. Methods

Device fabrication

We first fabricated a thin PDMS substrate film. The PDMS base and curing agent (Dow Corning, SYLGARD 184 silicone elastomer) were mixed in a ratio of 10:1 (w/w), after which the mixture was spin-coated on a plate before being transferred to an oven for curing at 70 °C for 2 hours. To prepare the graphite/PDMS composite, 10 wt% graphite (Sigma-Aldrich, 282863) was fully mixed with uncured PDMS (base: curing agent = 10:1, w/w) to form a black composite, enhancing image contrast. Toluene was added to the mixture to assist uniform dispersion and adjust its viscosity. The mixture was spin-coated on the transparent PDMS film (300 rpm, 60 s) to ensure uniformity. The film was transferred to an oven at 70 °C for 4 hours.

After preparing double-layer elastomer film, the micro trenches were fabricated with high-precision ultraviolet laser cutting. As shown in Figure S4b, the multilayer film was first placed on a glass plate, after which the film was patterned with the cross-shaped trench structure using a computer-controlled laser cutter.

Experimental set-up

The bespoke experimental setup aims to repeatedly apply vertical forces onto the sensor with given displacement. The setup's actuation mechanism consists of a magnetic linear motor (Faulhaber LM038004001) and a 3-D printed cone-shaped probe (1 mm tip diameter). The probe driven by the linear motor enables applying a controllable and high-precision displacement against the sensor surface. The position sensor in the linear motor provides highly accurate displacement measurements with a precision of 1 μm.

Calibration of force-displacement relationship

We placed the targeted sensing point at the centre of the bespoke probe. A high-precision scale (MAXREFDES82) was placed beneath the VBTS, which measures the forces applied. Subsequently, the position of the probe was adjusted to a height where it just contacted the sensor body. By initializing the linear motor with a pre-set displacement, we were able to measure the change of forces with the scale. We measured the force and The sensing points located at the top left corner of the sensor were tested: (1,1), (2,1), (2,2), (3,1), (3,2), and (3,3). Hence, the rest sensing points can be estimated symmetrically. For each sensing point, displacements of 0.5, 1.0, 1.5, and 1.75 mm were applied. And each force application was repeated for 10 seconds with a frequency of 0.5 Hz.

Acknowledgments

The authors would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council, UK. Grant number: EP/P012779/1. The authors would like to thank Prof Andrew Holmes for providing the laser processing equipment, and the Imperial College Research Computing Services team for promptly support in high performance computing.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] S. Pyo, J. Lee, K. Bae, S. Sim, and J. Kim, “Recent Progress in Flexible Tactile Sensors for Human-Interactive Systems: From Sensors to Advanced Applications,” *Advanced Materials*, vol. 33, no. 47, Nov. 2021, doi: 10.1002/adma.202005902.
- [2] S. Chun *et al.*, “An artificial neural tactile sensing system,” *Nat Electron*, vol. 4, no. 6, pp. 429–438, Jun. 2021, doi: 10.1038/s41928-021-00585-x.
- [3] J. C. Yang, J. Mun, S. Y. Kwon, S. Park, Z. Bao, and S. Park, “Electronic Skin: Recent Progress and Future Prospects for Skin-Attachable Devices for Health Monitoring, Robotics, and Prosthetics,” *Advanced Materials*, vol. 31, no. 48, Nov. 2019, doi: 10.1002/adma.201904765.
- [4] M. Shi *et al.*, “Self-Powered Analogue Smart Skin,” *ACS Nano*, vol. 10, no. 4, pp. 4083–4091, Apr. 2016, doi: 10.1021/acsnano.5b07074.
- [5] W. Gao *et al.*, “Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis,” *Nature*, vol. 529, no. 7587, pp. 509–514, Jan. 2016, doi: 10.1038/nature16521.
- [6] M. Shi, H. Wu, J. Zhang, M. Han, B. Meng, and H. Zhang, “Self-powered wireless smart patch for healthcare monitoring,” *Nano Energy*, vol. 32, pp. 479–487, Feb. 2017, doi: 10.1016/j.nanoen.2017.01.008.
- [7] L. E. Osborn *et al.*, “Prosthesis with neuromorphic multilayered e-dermis perceives touch and pain,” *Sci Robot*, vol. 3, no. 19, Jun. 2018, doi: 10.1126/scirobotics.aat3818.
- [8] A. P. Gerratt, H. O. Michaud, and S. P. Lacour, “Elastomeric electronic skin for prosthetic tactile sensation,” *Adv Funct Mater*, vol. 25, no. 15, pp. 2287–2295, Apr. 2015, doi: 10.1002/adfm.201404365.
- [9] H. Sun, K. J. Kuchenbecker, and G. Martius, “A soft thumb-sized vision-based sensor with accurate all-round force perception,” *Nat Mach Intell*, vol. 4, no. 2, pp. 135–145, Feb. 2022, doi: 10.1038/s42256-021-00439-3.
- [10] C. Xu, S. A. Solomon, and W. Gao, “Artificial intelligence-powered electronic skin,” *Nat Mach Intell*, vol. 5, no. 12, pp. 1344–1355, Dec. 2023, doi: 10.1038/s42256-023-00760-z.
- [11] S. Zhang *et al.*, “Hardware Technology of Vision-Based Tactile Sensor: A Review,” *IEEE Sens J*, vol. 22, no. 22, pp. 21410–21427, Nov. 2022, doi: 10.1109/JSEN.2022.3210210.
- [12] L. Van Duong and V. A. Ho, “Large-Scale Vision-Based Tactile Sensing for Robot Links: Design, Modeling, and Evaluation,” *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 390–403, Apr. 2021, doi: 10.1109/TRO.2020.3031251.
- [13] B. Ward-Cherrier *et al.*, “The TacTip Family: Soft Optical Tactile Sensors with 3D-Printed Biomimetic Morphologies,” *Soft Robot*, vol. 5, no. 2, pp. 216–227, Apr. 2018, doi: 10.1089/soro.2017.0052.
- [14] N. J. Ferrier and R. W. Brockett, “Reconstructing the Shape of a Deformable Membrane from Image Data,” *Int J Rob Res*, vol. 19, no. 9, pp. 795–816, Sep. 2000, doi: 10.1177/02783640022067184.
- [15] K. Kamiyama, K. Vlack, T. Mizota, H. Kajimoto, K. Kawakami, and S. Tachi, “Vision-based sensor for real-time measuring of surface traction fields,” *IEEE Comput Graph Appl*, vol. 25, no. 1, pp. 68–75, Jan. 2005, doi: 10.1109/MCG.2005.27.
- [16] K. Sato, K. Kamiyama, N. Kawakami, and S. Tachi, “Finger-Shaped GelForce: Sensor for Measuring Surface Traction Fields for Robotic Hand,” *IEEE Trans Haptics*, vol. 3, no. 1, pp. 37–47, Jan. 2010, doi: 10.1109/TOH.2009.47.
- [17] D. F. Gomes, Z. Lin, and S. Luo, “GelTip: A Finger-shaped Optical Tactile Sensor for Robotic Manipulation,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Oct. 2020, pp. 9903–9909. doi: 10.1109/IROS45743.2020.9340881.
- [18] W. Yuan, S. Dong, and E. Adelson, “GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force,” *Sensors*, vol. 17, no. 12, p. 2762, Nov. 2017, doi: 10.3390/s17122762.
- [19] M. K. Johnson and E. H. Adelson, “Retrographic sensing for the measurement of surface texture and shape,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2009, pp. 1070–1077. doi: 10.1109/CVPR.2009.5206534.
- [20] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson, “Microgeometry capture using an elastomeric sensor,” *ACM Trans Graph*, vol. 30, no. 4, pp. 1–8, Jul. 2011, doi: 10.1145/2010324.1964941.

- [21] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, “Measurement of shear and slip with a GelSight tactile sensor,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, May 2015, pp. 304–311. doi: 10.1109/ICRA.2015.7139016.
- [22] C. Trueeb, C. Sferrazza, and R. D’Andrea, “Towards vision-based robotic skins: a data-driven, multi-camera tactile sensor,” in *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*, IEEE, May 2020, pp. 333–338. doi: 10.1109/RoboSoft48309.2020.9116060.
- [23] C. Sferrazza and R. D’Andrea, “Transfer learning for vision-based tactile sensing,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Nov. 2019, pp. 7961–7967. doi: 10.1109/IROS40897.2019.8967571.
- [24] N. Funk, E. Helmut, G. Chalvatzaki, R. Calandra, and J. Peters, “Evetac: An Event-Based Optical Tactile Sensor for Robotic Manipulation,” *IEEE Transactions on Robotics*, vol. 40, pp. 3812–3832, 2024, doi: 10.1109/TRO.2024.3428430.
- [25] L. Fu, D. K. Saha, S. Shere, Y. Li, and H. Godaba, “ELTac: A Vision-Based Electroluminescent Tactile Sensing Skin for Force Localization and Magnitude Estimation,” *IEEE Sens J*, vol. 24, no. 5, pp. 6846–6855, Mar. 2024, doi: 10.1109/JSEN.2024.3354290.
- [26] M. Shi, Y. Zhang, X. Guo, and E. Yeatman, “Microstructure-Enhanced Vision-Based Tactile Sensor,” in *2023 22nd International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers)*, Kyoto, Japan: IEEE, Jun. 2023.

Supplementary Information

High-Sensitivity Vision-Based Tactile Sensing Enhanced by Microstructures and Lightweight CNN

Mayue Shi^{1*}, Yongqi Zhang¹, Xiaotong Guo¹ and Eric M. Yeatman¹

¹Department of Electrical and Electronic Engineering, Imperial College London, London, UK.

*Email: m.shi16@imperial.ac.uk.

Table of Contents

Supplementary Note 1 Mathematical Model and Analytical Approach	2
References	4
Supplementary Figure S3	5
Supplementary Figure S4	6
Supplementary Figure S5	7
Supplementary Figure S6	8
Supplementary Table S1	9

Supplementary Note 1 | Mathematical Model and Analytical Approach

A comprehensive mathematical model to analyse the bending behaviour of a thin PDMS film with fixed boundaries and uniformly distributed V-shaped notches. This approach simplifies the complex behaviour of the sensor film by representing it as an array of fixed-fixed beams. Each beam bends independently in response to an applied load, assuming that cross-sections remain undeformed during bending.

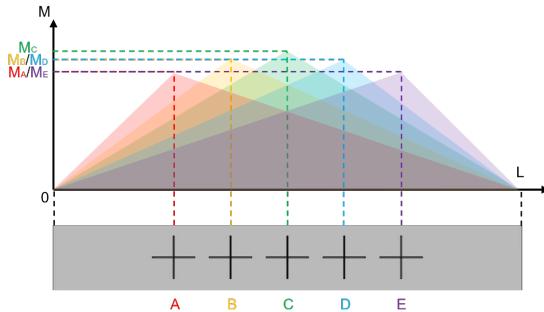


Figure S1: Comprehensive moment diagram of the system, taking into consideration the loads applied at notched locations A, B, C, D, and E. The moments for each location are denoted correspondingly as M_A , M_B , M_C , M_D , and M_E .

To analyse the bending behaviour, we utilised the flexibility method, a powerful tool for addressing the complexities of statically indeterminate systems. By simplifying the third-degree statically indeterminate system to a second-degree one, we focused on shear and moment reactions, neglecting the axial reaction. This method involves releasing one of the fixed ends to introduce degrees of freedom associated with translation and rotation, making the analysis more manageable. We can determine the redundant reaction forces and construct detailed moment diagrams illustrating the bending of fixed-fixed beams with notches, as demonstrated in **Figure S1**.

Saint-Venant's principle [1], a fundamental concept in structural engineering, asserts that the precise distribution of a load becomes less critical as we move away from the loaded region, as long as the overall load resultants remain accurate. In our specific scenario, we can apply Saint-Venant's principle to infer that stresses on a boundary considerably distant from the applied load remain relatively unaffected. The alterations in stress and strain primarily occur in the vicinity of the load application regions. Given that the notches are positioned on the opposite side from where the load is applied and considering the thinness of the notches relative to the beam's thickness, we can confidently employ Saint-Venant's Principle in our analysis. To simplify our analysis leveraging Saint-Venant's principle, we consider the regions immediately surrounding the notch as stress-free areas. We start from the central location (Location C), which proves to be the most convenient for analysis due to its symmetrical properties.

Here we represent the stress-free area as an isosceles triangle with a height of α and a base length of 2α , as shown in Figure 3.2a. This approach simplifies computations and aligns with Saint-Venant's principle's application.

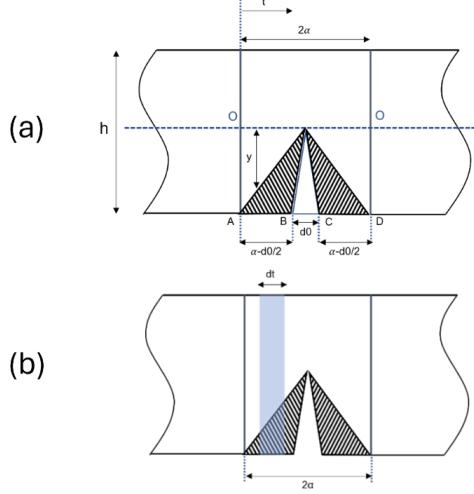


Figure S2. (a) Simplifying the analysis of the notched beam utilizing Saint-Venant's Principle at the central location C. (b) When integrating over the entire expanded area, represented by an isosceles triangle, we can obtain the deformation of the notch width. This integration process allows us to quantify the extent of deformation experienced by the notch width during the beam's bending.

In this context, d_0 represents the original notch width situated at the centre of the expanded configuration. According to Saint-Venant's principle, the expanded sections (shaded areas) will experience negligible deformation during the beam's bending. Consequently, the extension in the expanded configuration (segment \overline{AD}) is equal to the extension in the original configuration (segment \overline{BC}), denoted as Δd .

Extending from the geometric and physical relationships, we derived the following expressions to delineate this relationship.

$$\Delta d_c = \int_0^{2\alpha} \epsilon dt = 2 \int_0^\alpha \epsilon dt \quad (1)$$

$$\Rightarrow \Delta d_c = \frac{3M_c}{wE\alpha} \quad (2)$$

Here, apart from the parameters defined previously, ϵ denotes the strain from the flexural equation $\epsilon E = \frac{My}{I}$, and E denotes the effective Young's Modulus of the composite material, determined through experimental calibration.

When analysing intermediate locations (A, B, D, and E) that are off-centre, a notable asymmetry becomes evident regarding the stress-free areas. These stress-free areas closely resemble the shape of the respective moment diagrams at each location. Saint-Venant's principle indicates that the stress-free area is inversely related to the deformation, where smaller stress-free areas experience more significant deformations and vice versa. Therefore, during integration similar to the central location, we should integrate using a shape that is a horizontal mirror image of the corresponding moment diagram to accurately capture the varying deformation across these off-central locations. This observation emphasizes the importance of considering the shape of stress-free areas in accurately analysing and interpreting the beam's behaviour at different positions. Take one intermediate location *A* as an example, the deformations at the left and right edges of the notch can be expressed as:

$$\Delta d_A = \Delta d_{A_1} + \Delta d_{A_2} = \frac{49M_A}{22wE\alpha} + \frac{17M_A}{22wE\alpha} \quad (3)$$

References

- [1] R. v. Mises, "On Saint Venant's principle," *Bulletin of the American Mathematical Society*, vol. 51, no. 8, pp. 555–562, 1945, doi: 10.1090/S0002-9904-1945-08394-3.

Supplementary Figure S3

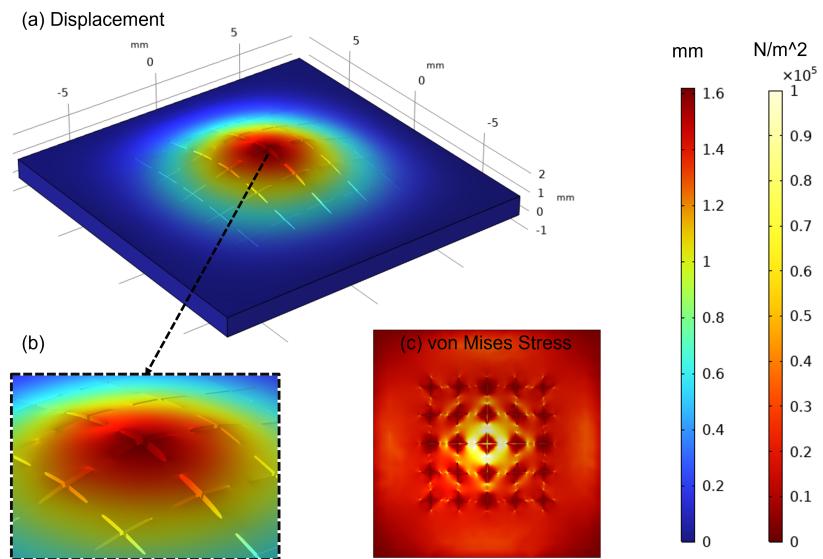


Figure S3: Finite element analysis of (a) deformation and (b) stress distribution with a force of 60 mN applied at the bottom centre (3, 3) of the sensor along the y-direction (diameter of contact area: 1 mm).

Supplementary Figure S4

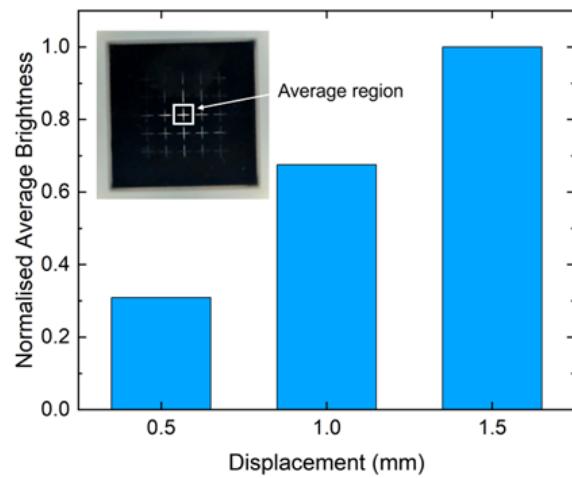


Figure S4: Variation of average brightness with different displacements. The contact point was at (3,3), the square average region fully included the central cross.

Supplementary Figure S5

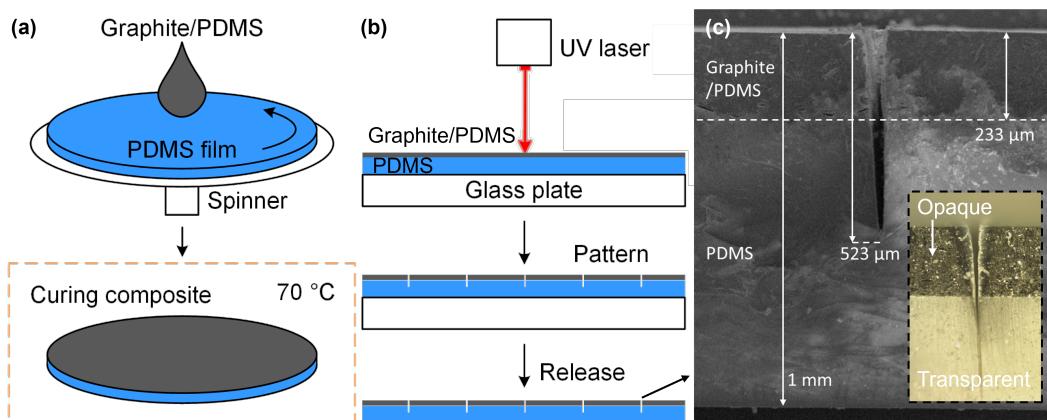


Figure S5: Fabrication processes and characterisation. (a) preparation of the film with the spinning coating process. (b) micromachining with high-precision laser cutting. (c) Characterisation of a micro trench with SEM and optical microscope.

Supplementary Figure S6

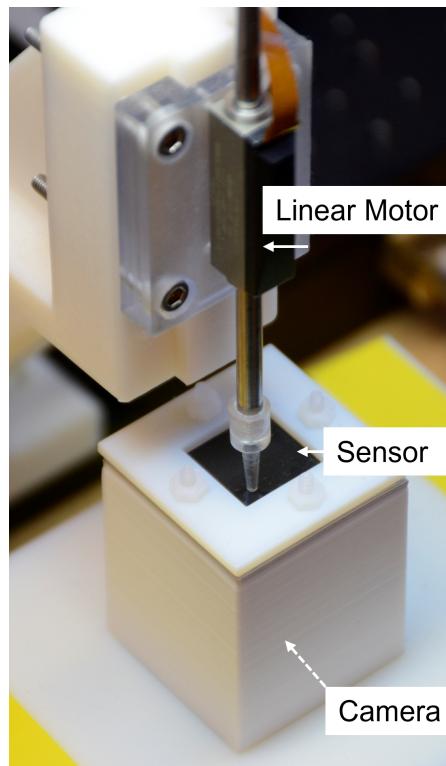


Figure S6: Experimental set-up. The setup includes a linear motor (Faulhaber LM083004001), a sensor, and a camera. Diameter of contact area: 1 mm.

Supplementary

Table S1

Table S1: Metrics for evaluating model performance

Metrics		z, e.g. displacement			
Model		MSE	RMSE	MAE	R ²
CNN_1		0.002225	0.047166	0.030523	0.988868
CNN_3		0.002045	0.045216	0.027137	0.989769
CNN_5		0.002002	0.044741	0.026321	0.989983
CNN_7		0.001955	0.04422	0.025705	0.990215
Metrics		x-coordinate			
Model		MSE	RMSE	MAE	R ²
CNN_1		0.000553	0.023523	0.018525	0.999723
CNN_3		0.00004	0.006314	0.004997	0.99998
CNN_5		0.000021	0.004588	0.003631	0.999989
CNN_7		0.000007	0.002654	0.00205	0.999996
Metrics		y-coordinate			
Model		MSE	RMSE	MAE	R ²
CNN_1		0.000873	0.029544	0.022877	0.999564
CNN_3		0.000078	0.008838	0.006963	0.999961
CNN_5		0.000017	0.004092	0.003249	0.999992
CNN_7		0.000006	0.002523	0.001991	0.999997