



OPEN

A soft thumb-sized vision-based sensor with accurate all-round force perception

Huanbo Sun ¹✉, **Katherine J. Kuchenbecker** ² and **Georg Martius** ¹✉

Vision-based haptic sensors have emerged as a promising approach to robotic touch due to affordable high-resolution cameras and successful computer vision techniques; however, their physical design and the information they provide do not yet meet the requirements of real applications. We present a robust, soft, low-cost, vision-based, thumb-sized three-dimensional haptic sensor named Insight, which continually provides a directional force-distribution map over its entire conical sensing surface. Constructed around an internal monocular camera, the sensor has only a single layer of elastomer over-moulded on a stiff frame to guarantee sensitivity, robustness and soft contact. Furthermore, Insight uniquely combines photometric stereo and structured light using a collimator to detect the three-dimensional deformation of its easily replaceable flexible outer shell. The force information is inferred by a deep neural network that maps images to the spatial distribution of three-dimensional contact force (normal and shear). Insight has an overall spatial resolution of 0.4 mm, a force magnitude accuracy of around 0.03 N and a force direction accuracy of around five degrees over a range of 0.03–2 N for numerous distinct contacts with varying contact area. The presented hardware and software design concepts can be transferred to a wide variety of robot parts.

Robots have the potential to perform useful physical tasks in a wide range of application areas^{1–4}. To robustly manipulate objects in complex and changing environments, a robot must be able to perceive when, where and how its body is contacting other things. Although widely studied and highly successful for environment perception at a distance, centrally mounted cameras and computer vision are poorly suited to real-world robot contact perception due to occlusion and the small scale of the deformations involved. Robots instead need touch-sensitive skin, but few haptic sensors exist that are suitable for practical applications.

Recent developments have shown that machine-learning-based approaches are especially promising for creating dexterous robots^{5,6}. In such self-learning scenarios and real-world applications, the need for extensive data makes it particularly critical that sensors are robust and keep providing good readings over thousands of hours of rough interaction. Importantly, machine learning also opens new possibilities for tackling this haptic sensing challenge by replacing handcrafted numeric calibration procedures with end-to-end mappings learned from data⁷.

Many researchers have created haptic sensors⁸ that can quantify contact across a robot's surfaces: previous successful designs produced measurements using resistive^{9–13}, capacitive^{14–16}, ferroelectric¹⁷, triboelectric¹⁸ and optoresistive^{19,20} transduction approaches. More recently, vision-based haptic sensors^{21–26} have demonstrated a new family of solutions, typically using an internal camera that views the soft contact surface from within; however, these existing sensors tend to be fragile, bulky, insensitive, inaccurate and/or expensive. By considering the goals and constraints from a fresh perspective, we have invented a vision-based sensor that overcomes these challenges and is thus suitable for robotic dexterous manipulation.

Table 1 provides a detailed comparison of representative state-of-the-art sensors. We highlight the most important differences and refer the reader to the Methods for a more thorough examination. The mechanical designs of all previous sensors employ multiple functional layers, which are complex to fabricate and can be

delicate. Insight is the only sensor with a single soft layer. Many tasks benefit from a large three-dimensional sensing surface rather than small two-dimensional sensing patches; however, only a few other sensors offer three-dimensional surfaces^{25,27–29}. Some of them require special lenses²⁵ or use multiple cameras²⁷, whereas others are more fragile^{28,29}. Insight needs only a single camera and simple manufacturing techniques. Depending on their mechanical design, sensors also have widely varying sensing surface area and sensor volume. We provide area per volume (A/V) in Table 1 as a measure of compactness and find that Insight is among the most compact vision-based sensors with the largest sensing surface.

Most existing sensors provide only localization of a single contact^{20,25,27,28,30}; some also provide a force magnitude^{9,23,31} without force direction. Others are specialized for measuring contact area shape^{21,29,32}. Although real contacts will be multiple and complex, a spatially extended map of three-dimensional contact forces over the surface, which we call a force map, is only rarely provided (for example, ref. ²²). Insight is the only sensor that provides a force map across a three-dimensional surface such that a robot can have detailed directional information about simultaneous contacts. Many sensors rely on analytical data processing^{22,25,28,33}, which requires careful calibration; it is difficult to obtain correct force amplitudes with such an approach as materials are often inhomogeneous and the assumption of linearity between deformation and force is often violated. Data-driven approaches such as those used with a BioTac⁹, GelSight²¹, OmniTact²⁷ and Insight can deal with these problems but require copious quality data.

This paper presents a new soft thumb-sized sensor with all-round force-sensing capabilities enabled by vision and machine learning; it is durable, compact, sensitive, accurate and affordable (less than \$100). As it consists of a flexible shell around a vision sensor, we name it Insight. Although initially designed for dexterous manipulation and behavioural learning, our sensor is suitable for many other applications and our technology can be adapted to create a variety of three-dimensional haptic sensing systems.

¹Autonomous Learning Group, Max Planck Institute for Intelligent Systems, Tübingen, Germany. ²Haptic Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, Germany. ✉e-mail: huanbo.sun@tuebingen.mpg.de; georg.martius@tuebingen.mpg.de

Table 1 | A comparison between state-of-the-art haptic sensors and our design (Insight, highlighted in bold). An upward (downward) arrow indicates that higher (lower) values are better

Sensor	Transduction Method	No. of Layers ↓	Surface Shape ↑	Area ↑ (mm ²)	A/V ↑ (mm ⁻¹)	Data Processing	Output Format	Sensing error ↓				Notes
								P(mm)	F _n (N)	F _s (N)	α(°)	
BioTac	Resistive	2	Half 3D	484	0.060	FCN	Location + force	1.4 (ref. ⁵⁹)	0.85	-0.48 (ref. ⁹)	10 ⁶⁰	Costly and delicate
Lee et al. ³⁰	Capacitive	5	2D	484	1.000	—	Location	2	—	—	—	10 mN Full-scale range
Yan et al. ³¹	Hall effect	2	2D	324	0.179	FCN	Location + force	0.1	0.15	—	—	Blind to shear force
SLIMS ²⁰	Optical	3	1D	350	0.050	—	Location	10	—	—	—	Sensitive to ambient light
GelSight ²¹	Camera + PS + markers	4	2D	250	0.003	CNN	Shape + force	—	0.67	-0.17	—	Complex to manufacture
GelSlim ²²	Camera + PS + markers	5	2D	1,200	0.006	iFEM	Force map	—	0.32	-0.22	—	Tears after 1,500 contacts
OmniTact ²⁷	Five cameras + PS	2	3D	3,110	0.083	ResNet	Location	0.4	—	—	—	Five cameras, ambient light
GelTip ²⁸	Camera + PS	2	3D	2,513	0.052	Numeric	Location	5	—	—	—	Imaging artefacts on tip
TacTip ²⁵	Camera + markers	3	3D	2,500	0.025	Numeric	Location	0.2	—	—	—	Marker density limits resolution
C.S. & R.D'A. ²³	Camera + beads	3	2D	900	0.008	FCN	Location + force	0.2	0.05	—	—	Heavy, blind to shear
C.S., T.B. & R.D'A. ⁵²	Camera + beads	3	2D	900	0.008	CNN	Force map	—	0.13	0.05	—	No real shear experiment
DIGIT ³²	Camera + PS + markers	3	2D	304	0.031	ResNet	Shape	—	—	—	—	Difficult to extend to 3D
Romero et al. ²⁹	Camera + PS	2	Half 3D	2,069	0.084	Numeric	Shape	—	—	—	—	Limited sensing area
Insight	Camera + PS + SL	1	3D	4,800	0.088	ResNet	Force map	0.4	0.03	0.03	5	This work

CNN, convolutional neural network; FCN, fully connected neural network; iFEM, inverse finite element method; PS, photometric stereo; SL, structured light.

Figure 1 shows the principles behind the design of Insight. The skin is made of a soft elastomer over-moulded³⁴ on a hollow stiff skeleton to maintain the sensor's shape and allow for high interaction forces without damage (Fig. 1b). It utilizes shading effects³⁵ and structured light³⁶ to monitor the three-dimensional deformation of the sensing surface with a single camera from the inside (Fig. 1c). The sensor's output is computed by a data-driven machine-learning approach^{10,12}, which directly infers distributed contact-force information from raw camera readings, avoiding complicated calibration or any handcrafted post-processing steps (Fig. 1d).

We evaluate Insight against several rigorous performance criteria. When indented by a hemispherical tip with a diameter of 4 mm at a force amplitude of up to 2 N, the sensor can achieve an average localization accuracy and force accuracy of around 0.4 mm and 0.03 N, respectively. By directly estimating both the normal and shear components of each applied force vector, the sensor reaches an average directional estimation error of around 5°. Moreover, in the absence of contact, Insight is sensitive enough to recognize its posture relative to gravity based only on the deformations caused by its own self-weight, which are not detectable by the human eye.

Principles of operation and design

At the core of our design is a single camera that observes the sensor's opaque over-moulded elastic shell from the inside (Fig. 1a). Photometric effects and structured lighting enable it to detect the tiny deformations of the sensor surface that are caused by

physical contact. Instead of computing the contact force vectors numerically from the observed deformations using elastic theory^{33,37}, which poses limiting assumptions, we use a machine-learning approach that translates images directly to force distribution maps. The details are shown in Fig. 1 and explained below.

Mechanics. We aim at a compliant and sensitive sensing surface due to the favourable properties of soft materials for manipulating objects³⁸, for safer interactions around humans³⁹, and to limit the instantaneous impact forces during unforeseen collisions⁴⁰. Nevertheless, soft materials alone cannot withstand large interaction forces and are deformed by gravity and inertial effects⁴¹.

To ensure a compliant sensing surface, high contact sensitivity and robustness against self-motion, we design a soft-stiff hybrid structure using over-moulding (Fig. 1b)³⁴. The structure is composed of two parts: a flexible elastomer to sense contact and an aluminium skeleton to support the sensing surface. The resulting sensor is not only sufficiently structurally stable to keep its overall shape under high contact forces, but it is also sensitive enough for gentle interaction forces to cause local deformations. By contrast to our approach, other successful curved vision-based sensors such as GelTip²⁸ and that in Romero and colleagues²⁹ solve the stability problem with a smooth and uniform support structure out of transparent glass, acrylic or resin²⁹, allowing for good imaging quality and acting as a light guide. Our metal skeleton can be designed independent of the lighting.

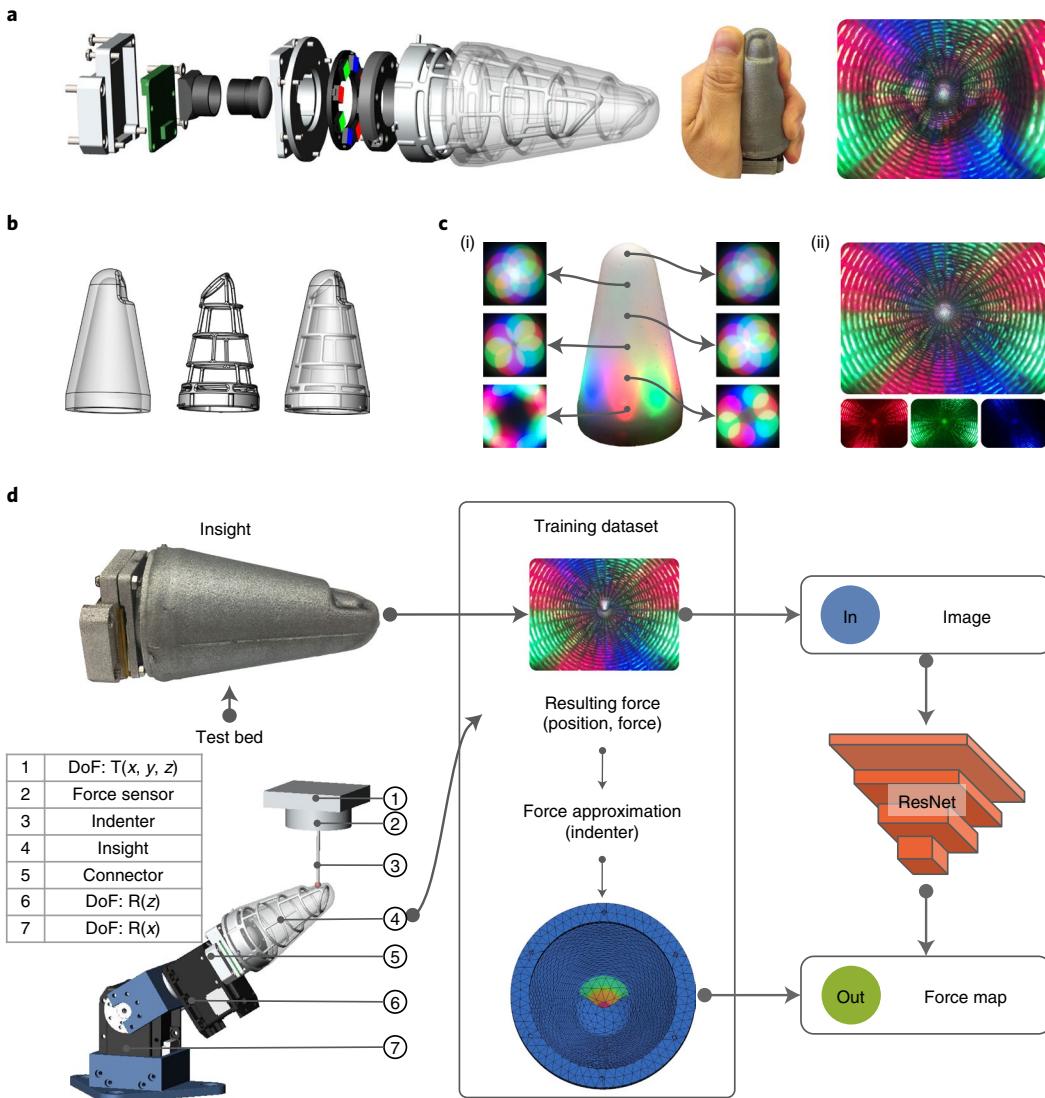
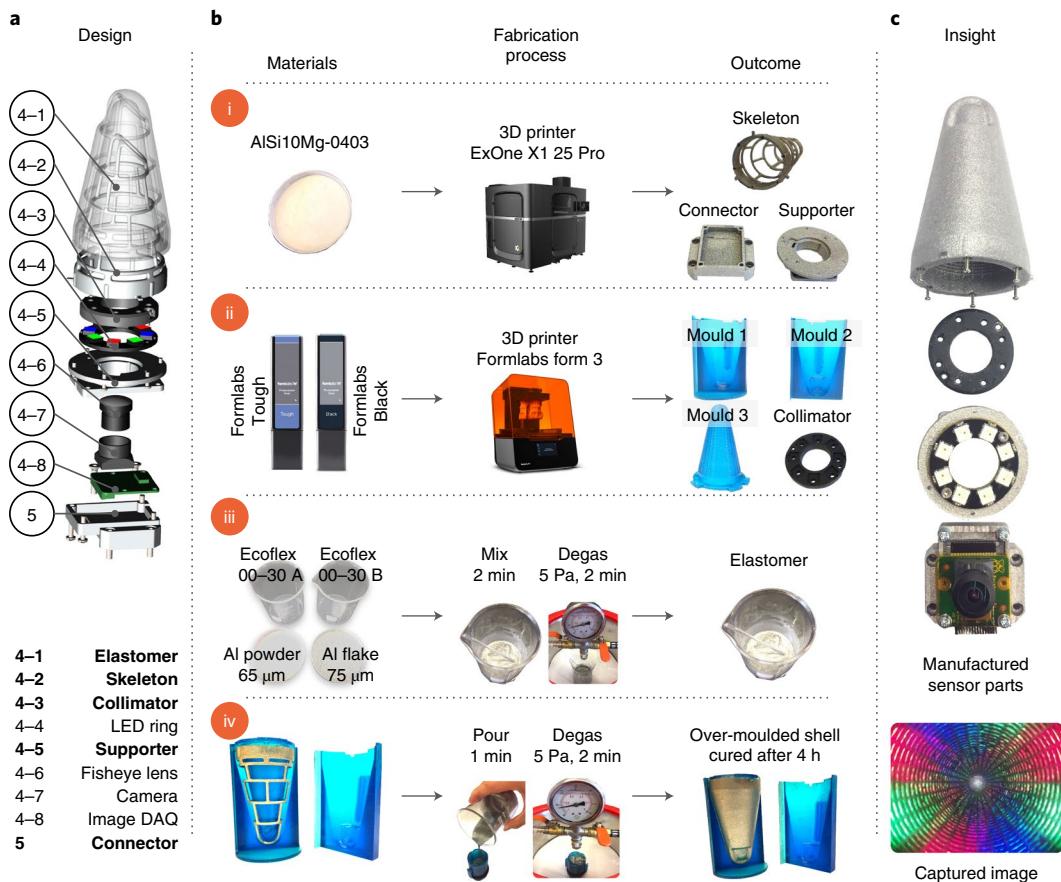


Fig. 1 | The design principles of Insight. **a**, The overall structure of the sensor with its hybrid mechanical construction and internal imaging system. For comparison, the sensor is shown in a human hand next to the corresponding camera view. **b**, The pure elastomer (left), the stiff hollow skeleton (middle) and both over-moulded together (right). **ci**, The internal lighting using a translucent shell: the LED ring with apertures creates light cones, visualized by their projections on flat horizontal planes. **cii**, The light projection patterns within, as seen by the camera in the undeformed opaque shell. **d**, The data processing pipeline. The machine-learning model is trained on data collected by an automatic test bed; T and R stand for translation and rotational test-bed movements, respectively. Each data point combines one image from the camera with the indenter's contact position and orientation, contact force vector and diameter, which are used to calculate a ground-truth force distribution map from an approximate model under consistent contact forces.

Insight's shell is hollow so that the entire system is lightweight; avoiding direct contact with any optical elements (by contrast to ref. ²⁷) also reduces the chances of image distortion and system damage. Constructing a single elastomer layer that serves all purposes is a simple, compact, robust and wireless solution for haptic sensing. All other vision-based haptic sensors are built from multiple layers of different materials (for example, protective coatings, marker sheets, elastomers, adhesive sheets). Durability issues often arise due to non-permanent attachment between layers (for example, refs. ^{21,27,28,42}). Another design consideration is the opaqueness of the elastomer. Our elastomer layer is opaque enough to block all interference from ambient light, ensuring reliable output even under bright lighting conditions. Sensors that have a thin opaque coating on top of transparent elastomer may struggle to achieve this property and/or maintain it over long-term use. To demonstrate the sensitivity of our approach, we include a thin, flat area of elastomer near the

sensor's end for higher-resolution perception of detailed shapes (akin to a tactile fovea).

Imaging. Two main techniques can be used to obtain three-dimensional information from a single camera. Photometric stereo³⁵ uses multiple images of the same scene with varying disparate light sources from different illumination directions to infer the three-dimensional shape from shading information. Structured light³⁶ is a single-shot three-dimensional surface-reconstruction technique that uses a unique light pattern and the fact that its appearance depends on the shape of the three-dimensional surface on which it is projected. Photometric stereo is generally better at capturing local details, whereas structured light is used for coarser global reconstruction^{43,44}. Photometric stereo is most effective when the illumination is nearly parallel to the surface, where the normal vectors of the deformed surface can be finely reconstructed from shading



information²¹. Sensors built on photometric stereo^{21,28,42} employ light guides to create this desirable tangential surface illumination, which is challenging for highly curved sensing surfaces^{28,29}. Structured light allows for more perpendicular lighting of the surface and improves with larger disparity between the light source and camera.

Insight is unique among haptic sensors as it combines photometric stereo and structured light to detect the deformation of a full three-dimensional cone-shaped surface in the single-camera single-image setting. Light-emitting diode (LED) sources around the camera produce distinct light cones (eight in our prototype, as shown in Fig. 1c). The lighting direction is adjusted through a collimator to introduce a suitable structured light pattern that favours locally parallel lighting for photometric stereo, as depicted in Extended Data Fig. 1c. In contrast to the light guides of other sensors^{21,28,29,42}, our collimator allows for flexible lighting and is independent of the support structure. When an area of the sensor surface is contacted from the outside, the surface orientation changes, which causes a difference in colour intensity through shading. The surface displacement also changes the distance of the surface to the camera, which can be detected with structured light cones due to the colour change per pixel.

Haptic information. Sensors can capture many types of haptic information such as vibration⁴⁵, deformation^{12,21,46}, undirected pressure distribution¹⁰ and directional force distribution^{22,33}. For robotics applications, a directional force distribution is the preferred form of contact information, as it describes the location and size of each contact region, as well as the local loading in the normal and shear

directions⁴⁷. Our proposed sensor is designed to deliver precisely this type of contact information, that is, a three-dimensional directional force distribution over a three-dimensional conical sensing surface represented by a fine mesh of points, where each point has three mutually orthogonal force elements.

In a classical estimation chain, the force distribution is inferred from the surface displacement using a linear stiffness matrix based on elastic theory³⁷. This approach is poorly suited for our design (as discussed in the Methods) so we employ a data-driven method to estimate the force distribution directly from the raw image input using machine-learning techniques, namely an adapted ResNet⁴⁸, which is a favoured deep CNN architecture. To collect reference data to train the neural network, we built a position-controlled five-degrees-of-freedom (DOF) test bed with an indenter that probes the designed sensor. A six-DOF force-torque sensor (ATI Mini40) measures the force vector applied to the indenter so we can simultaneously record ground truth forces and corresponding images from the camera inside the sensor. The target force distribution map corresponding to each contact is computed by a simple spatial approximation using the known force vector, contact location, and indenter diameter. A subset of all data is used to train the machine-learning model. The entire process is illustrated in Fig. 1d and detailed in the Methods.

Fabrication. As depicted in Fig. 2, fabrication of Insight includes three main aspects: the imaging system, mechanical components and optical properties. An explanation of the design choices and further details of the fabrication process can be found in the Methods.

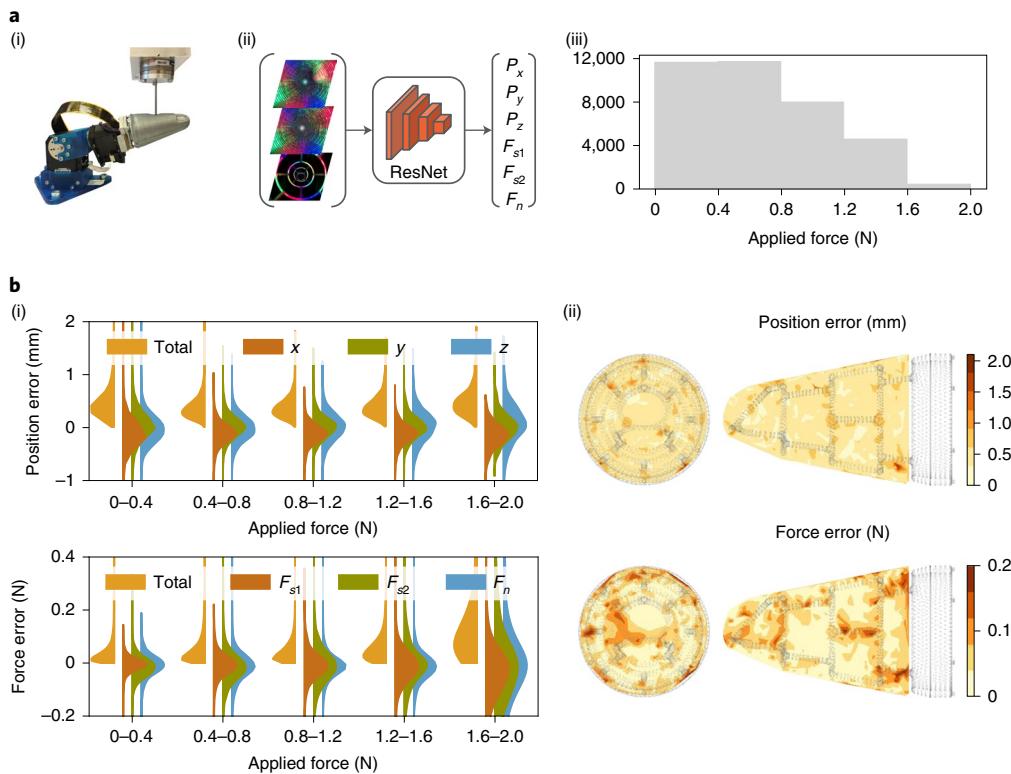


Fig. 3 | Single contact performance with direct estimation. **a**, The estimation pipeline for inferring single contact position and force. **a*i***, The real experimental set-up in which the test bed probes Insight and collects data. **a*ii***, The machine-learning model: the inputs are three images (raw image, reference image and skeleton image), and the outputs are the contact location (P_x , P_y and P_z denote the coordinates of the contact in the sensor's reference frame) and contact force vector (F_{s1} , F_{s2} and F_n). **a*iii***, A histogram of the forces applied in the data collection procedure. **b**, Statistical evaluation of the sensor's performance on the test data. **b*i***, The localization and force estimation performance grouped by applied force magnitude. The red-, green- and blue-coloured half-violins show the distribution of deviations in the x , y and z directions, respectively. The force is predicted relative to the surface in normal direction F_n and two shear directions F_{s1} and F_{s2} . The orange half-violins stand for the resulting total errors. **b*ii***, The spatial distribution of the localization and force quantification errors for the same test data.

Performance

The sensor's performance is evaluated with respect to both accuracy and sensitivity. The first measure of accuracy is direct single-contact estimation: a contact force needs to be localized, and its magnitude and direction must be inferred. Second is force distribution estimation for single contact: the contact area and directional force distribution over the entire sensing surface are inferred. We also provide qualitative results for multiple contacts. Finally, we evaluate Insight's sensitivity by studying whether it can perceive gravitational effects and characterizing its ability to detect shapes contacting the tactile fovea.

Accuracy of direct contact estimation. Our primary way to assess accuracy is to quantitatively evaluate the system's ability to localize contacts and measure the applied force. First, we use a hemispherically tipped indenter with a diameter of 4 mm to probe a large number of points distributed across Insight's sensing surface (Fig. 3ai). In this procedure, we collect the images under contact and the contact force vectors from the force sensor, as well as the position of each contact on the sensor's surface using our five-DOF test bed (Fig. 1d). The histogram of the applied forces in Fig. 3aiii shows that most contacts have magnitudes smaller than 1.6 N, as we set this value as the threshold of data collection to avoid damaging the sensor. We then train a machine-learning model (modified ResNet⁴⁸ structure) to infer the contact information. The inputs to the model are the image under contact, a static reference image without contact, and a static image of the stiff skeleton for inhomogeneous elasticity encoding (recorded before over-moulding in a dark environment).

The outputs are the three-dimensional coordinates of the contact in the sensor's reference frame and the three-dimensional force components expressed in the local surface coordinate frame, as depicted in Fig. 3aii. Details on data collection and machine learning are summarized in the Methods.

We evaluate the single-contact direct estimation accuracy of localization and force sensing for an applied force magnitude up to 2 N, as shown in Fig. 3bi. All reported numbers are for test contact points that do not appear in the training data. The overall median localization precision is around 0.4 mm, and the force magnitude precision is approximately 0.03 N in the normal and shear directions. The force direction is estimated with a precision of approximately 5°. Notice that the test bed has an overall position precision of 0.2 mm, and the force-torque sensor has a force precision of 0.01/0.01/0.02 N ($F_x/F_y/F_z$). Insight's accuracy in localization is remarkably stable over different force ranges, whereas the error in force amplitude slightly increases with higher interaction force. For strong applied forces (over 1.6 N), the force accuracy becomes worse, presumably as we have insufficient training data for this domain (histogram in Fig. 3aiii). Another explanation is that high forces occur most often at locations near the stiff frame (Fig. 3bii), which deforms only a little. There is no noticeable difference in the localization and force accuracy in the sensor frame's x , y and z directions.

We particularly evaluate Insight's accuracy at localizing test contact points, as shown in Fig. 3bii. The accuracy is stable across the entire surface, and higher errors appear near the stiff frame. Only areas near the camera show a systematic performance drop; as our

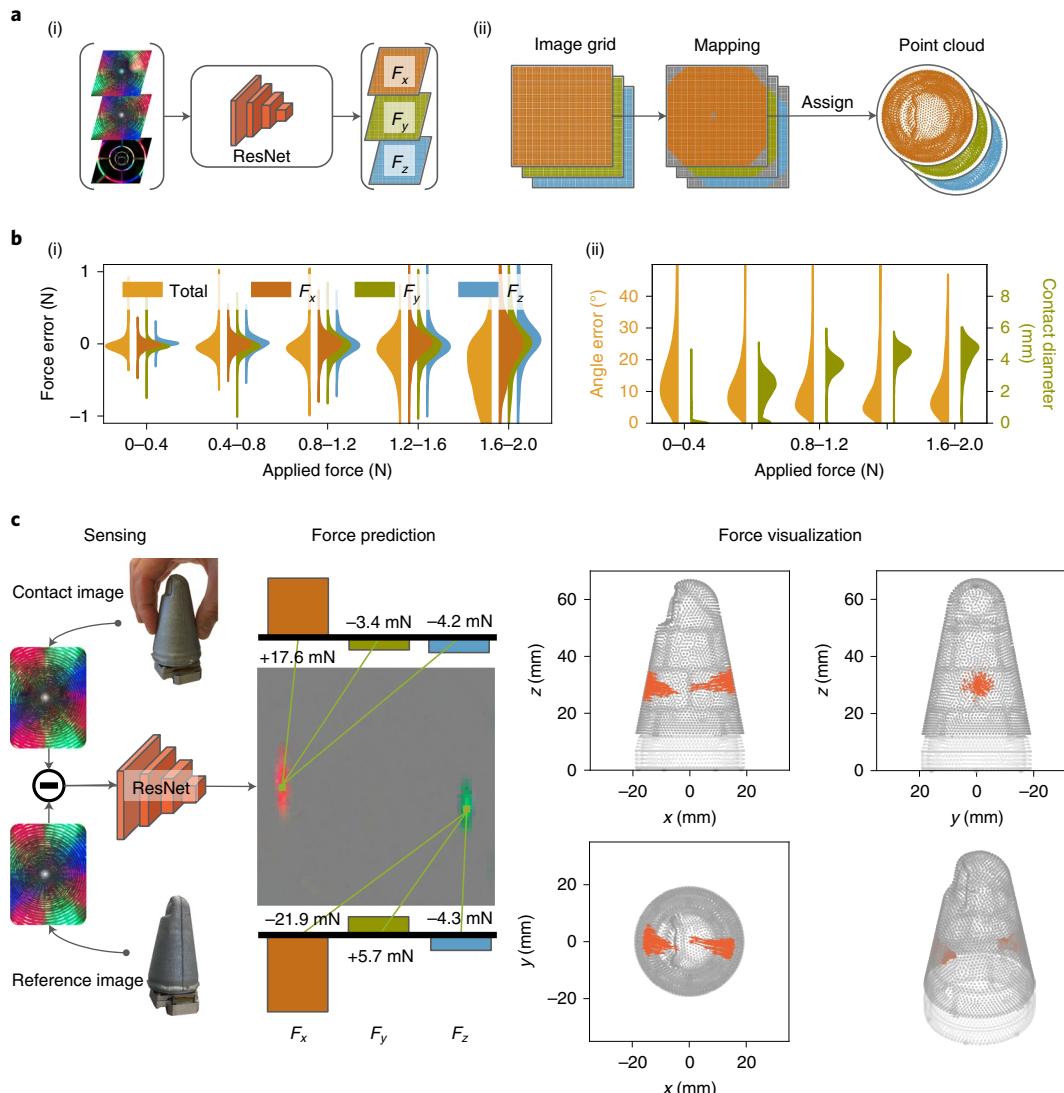


Fig. 4 | Performance evaluation of the force map. **a**, The pipeline of estimating the force distribution; the ResNet network transforms three images (raw, reference, skeleton) into the x - y - z force map image (**a**i), and its pixels are mapped to points on the sensing surface (**a**ii). **b**, The quantitative evaluation of the performance for force amplitude, force direction and contact area size inference grouped by applied force amplitude. **c**, The data flow and estimated force map when the sensor is pinched and rotated by two fingers.

camera has a 4:3 aspect ratio, it cannot see two opposite areas at the base of the shell, below the lowest ring of the stiff frame.

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

Insight has a three-dimensional curved surface and thus needs to output a force map with the same shape. We create a fine mesh of 3,800 points spanning the entire surface with an average spacing of 1 mm. Each point has three output values describing the force components it feels in the x , y and z directions expressed in the reference frame of the sensor.

Similar to the direct contact estimation, we also employ a machine-learning-driven pipeline. Instead of the six-dimensional output (Fig. 3aii), the network now produces the approximate force distribution map (Fig. 4ai) using only convolutional layers.

The map is estimated as a flat image with three channels (F_x , F_y , F_z) to describe nodal forces (the individual force on each point) in the x , y and z directions, respectively, mimicking the red, green and blue channels in a colourful image. Each pixel in the image corresponds to one point in the force map. The correspondence is established using the Hungarian assignment method⁴⁹, which minimizes the overall distance between pixels and points projected to the two-dimensional camera image, as shown in Fig. 4aii. Training the machine-learning model from collected data also requires target force distribution maps (Fig. 1d). As they are not measured directly, we approximate the force map applied by the indenter by distributing the measured total force locally across the surface. From a set of five diverse candidates, as detailed in Supplementary Section A.3, the approximation yielding the best performance in localization and force magnitude accuracy is selected (see Extended Data Fig. 2 and Supplementary Table 5).

The quantitative estimation accuracy for the force amplitude and direction is reported in Fig. 4b, grouped by force magnitude. The evaluation is based on the comparison between the three-dimensional force vectors summed across the predicted force map and the

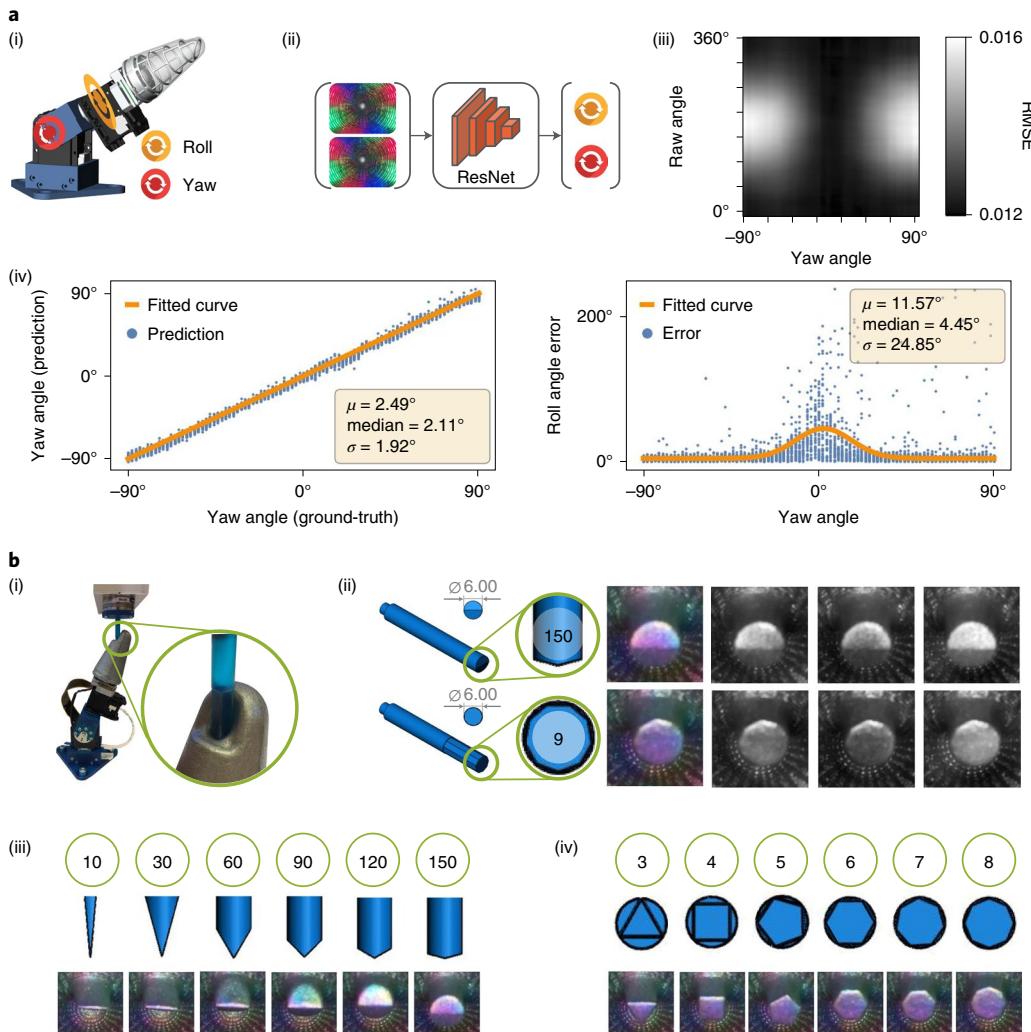


Fig. 5 | Evaluation of Insight's sensitivity to self-posture and detailed shapes. **a**, Quantitative evaluation of the sensor posture recognition. **a*i***, The experiment set-up (**i**) and the inference procedure for posture (roll and yaw angles) (**ii**). The network maps the difference between the current image and the reference image to posture coordinates. **a*iii***, The pixel-wise root-mean-square error (RMSE) of the image difference as the sensor is rotated to all possible roll and yaw angles. **a*iv***, A statistical summary of the posture estimation accuracy: the yaw angle estimation performance with a yellow fitted curve (left) and the roll angle estimation accuracy under different yaw angles (right). **b**, A qualitative evaluation of the tactile fovea for shape detection. **b*i***, The experimental set-up for applying differently shaped probes to the high-sensitivity region. **b*ii***, The system's perceptual limits for sharpness (v-shaped wedge with an included angle of 150°) and number of edges (nine-sided polygon) with an indenter diameter of 6 mm, along with their corresponding captured raw images and respective red, green and blue channels. **b*iii,iv***, The sharpness and edge tests: included angle and edge count (upper), indenter samples (middle) and captured images under indentation tests (lower).

ground-truth force vectors using the same single-contact dataset. The median error in inferring the total force is around 0.08 N, and the error grows with increasing force (Fig. 4*bi* and Supplementary Fig. 4). The system's tendency to slightly underestimate larger forces is probably caused by our force map approximation method, the influence of the skeleton, and the machine-learning method itself, which tends to estimate smooth force distributions rather than peaked maps. An ablation of the skeleton image as input leads to worse underestimation (Supplementary Section A.6), supporting part of this hypothesis. The median error in inferring the force direction is around 10° for low contact forces, and it decreases to 5° with higher applied forces (Fig. 4*bi*). Moreover, we can also localize the contact with a precision of around 0.6 mm based on the force map by averaging the locations of the 20 points with the highest force amplitudes (Supplementary Fig. 4). Supplementary Section A.6 analyses how this performance depends on the amount of data and the type of input provided to the network.

The contact area is estimated by identifying the points with predicted forces larger than 0.02 N. The diameter of this contact area increases with higher applied force and tends to overestimate by about 1 mm for a 4 mm indenter at high forces. Insight possesses a nail-shaped zone with a thinner elastomer layer (1.2 mm) and a sensing area of $13 \times 11 \text{ mm}^2$, as indicated in Fig. 5*b*. The median position and force errors in the tactile fovea are 0.3 mm and 0.026 N over an applied force range of 0.03–0.8 N, which shows better position accuracy and force accuracy than other sensing areas.

We use an indenter with 12 mm diameter to validate the force map inference performance and report details in Supplementary Fig. 5. The median position accuracy is 1 mm. For higher applied force, the underestimation of force magnitude is more pronounced. Force direction is measured to a high level of accuracy, achieving a median error of 8°. The median contact area estimates closely match expectations for a 12 mm indenter at each force level. As anticipated,

the predicted force map is inhomogeneous and shows higher forces near the skeleton.

Multiple simultaneous contacts. We also qualitatively demonstrate the sensor's performance during multiple complex contacts. Figure 4c shows the exemplary response to a human using two fingers to pinch and slightly twist the sensor. Each pixel of the force map contains the three force values estimated at that point. We facilitate interpretation by visualizing each contact force vector on the three-dimensional surface of the sensor. The experimenter's counter-clockwise twisting input can be seen in the slant of the force vectors when the sensor is viewed axially. Extended Data Fig. 3 and Supplementary Video 4 show the response for other contact situations. In our experiments, the sensor was consistently able to discriminate up to five simultaneous contact points and estimate each contact area in a visually accurate manner.

Sensitivity. The final two experiments evaluated Insight's sensitivity to subtle haptic stimuli. The sensor can accurately estimate its own orientation relative to gravity by visually observing the small gravity-induced deformations of the over-moulded elastomer (see Fig. 5a, Extended Data Fig. 4a and Supplementary Section A.3). Note that this experiment was conducted without any contacts and in a dark room to rule out other possible clues about self-posture. The median error for predicting yaw was 2.11°, and it was 4.45° for roll, with the highest errors for the roll angle around vertical, as expected. The camera was also found to capture relevant shape details when v-shaped wedges and extruded polygons were pressed into the tactile fovea (Fig. 5b).

Discussion

We present a soft haptic sensor named Insight that uses vision and learning to output a directional force map over its entire thumb-shaped surface. The sensor has a localization accuracy of 0.4 mm, force magnitude accuracy of 0.03 N and force direction accuracy of 5°. It can independently infer the locations, normal forces and shear forces of multiple simultaneous contacts—up to five regions in our evaluation. Moreover, the sensor is so sensitive that its quasi-static orientation relative to gravity can be inferred with an accuracy around 2°. A particularly sensitive tactile fovea with a thinner elastomer layer allows it to detect contact forces as low as 0.03 N and perceive the detailed shape of an object. A detailed comparison between Insight and other sensors can be found in Table 1 and the Methods.

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

The inhomogeneity of our sensor's surface might cause unwanted effects in some applications. Robotic systems that move with high angular velocities and high accelerations will probably see tactile sensing artefacts caused by inertial deformations of the soft sensing surface; data collected during dynamic trajectories could potentially mitigate these effects.

In general, our sensor design concept can be applied and extended to a wide variety of robot body parts with different shapes and precision requirements. The machine-learning architecture, training process and inference process are all general and can be applied to differently shaped sensors or other sensor designs. We also provide

ideas on how to adjust Insight's design parameters for other applications, such as the field of view of the camera, the arrangement of the light sources and the composition of the elastomer.

Methods

We conducted several experiments to make informed design choices and validate the functionality of Insight.

Sensor shape and camera view. The sensor is cone-shaped with a rounded tip to allow an all-round touch sensation in a structure similar to a human thumb. The sensor has a base diameter of 40 mm and a height of 70 mm. The Raspberry Pi camera v.2.0 (MakerHawk Raspberry Pi Camera Module 8 MP) has a resolution of 1,640 × 1,232 and a frame rate of 40 fps. With a 160° fisheye lens, the camera's field of view is 123.8° × 91.0°. See Supplementary Fig. 1 and Supplementary Table 1 for more details on the camera. Multiple cameras are recommended if the whole sensing surface cannot be seen by a single camera, as done in OmniTact²⁷, at the cost of increased wiring, material costs and computational load.

Light source and collimator. We use a commercial LED ring that contains eight tri-colour LEDs (WS2812 5050). The LED colours are programmed to be red (R), green (G), blue (B), R, G, R, B, G in circumferential order, and the relative brightnesses for the R, G and B light sources are 1:1:0.5, respectively. We designed a three-dimensionally printable collimator (three-dimensional printer, Formlabs Form 3; material, standard black; Formlabs owns the trademark and copyright of these names and pictures used in Fig. 2bii) with a tuned diameter (2.5 mm) and a radially tilted angle (3°) toward the outside to constrain the light-emitting path and create the structured light distribution (see Extended Data Fig. 1). Detailed analysis can be found in Supplementary Section A.1.

Soft surface material, skeleton and over-moulding. Insight's mechanical properties are optimized to ensure high sensitivity to contact forces, robustness against impact forces and low fatigue effects. We choose Smooth-On EcoFlex 00-30 silicone rubber as the mouldable material for the soft sensing surface because it is readily available and has a high elongation ratio of 900% (Supplementary Table 3). The skeleton is made of AlSi10Mg-0403 aluminium alloy, which can withstand forces up to 40 N in the shape of our prototype structure (Supplementary Table 4). These two materials are chosen based on their material data sheets and finite element analysis results³⁰. The elastomer is cast using three-dimensionally printed moulds (three-dimensional printer, Formlabs Form 3; material, tough; Formlabs owns the trademark and copyright of these names and pictures used in Fig. 2bii), and the skeleton is three-dimensionally printed in aluminium (three-dimensional printer, ExOne X1 25 Pro; material, AlSi10Mg-0403; ExOne GmbH owns the trademark and copyright of the name and picture used in Fig. 2bi). We combine the skeleton and the elastomer without adhesive by over-moulding, as described in Figs. 1b and 2b. Due to the working principle, the moulds require no special treatment; they are used straight from the three-dimensional printer, in contrast to, for example, Romero and colleagues²⁹. Furthermore, the manufacturing procedure is simple and requires only a single step.

The diameter of the skeleton beams and the thickness of the surrounding elastomer are optimized for robustness, as described in Extended Data Fig. 5. Finite element analysis revealed that the system's sensitivity to contact forces improved by positioning the skeleton not in the centre of the elastomer layer but closer to the inner surface.

Optical properties. We need a material with the right reflective properties (albedo, specularity) for the sensing surface. It should not be too reflective as reflections saturate the camera and diminish sensitivity. Simultaneously, no point on the surface should be very dark, as the camera needs to detect changes in reflected light. Moreover, the material has to prevent ambient light from perturbing the image. The sensing surface is made of a flexible and mouldable translucent elastomer mixed with aluminium powder and aluminium flakes. The aluminium powder makes the surface opaque to ambient light, and the aluminium flakes adjust the reflective properties, as shown in Figs. 1c and 2b, and Supplementary Fig. 2 and Supplementary Table 2. Aluminium powder with 65 µm particle diameter and aluminium flake with 75 µm particle diameter are mixed with the EcoFlex 00-30 in a weight ratio of 20:3:400 to ensure proper light reflection properties and opacity (see Supplementary Fig. 2 and Supplementary Table 2).

Finite element analysis. To analyze the over-moulded shell's mechanical properties during the design process, we built a finite element model using Ansys³¹. The model includes Young's modulus and Poisson's ratio for both EcoFlex 00-30 (70 kPa, 0.4999) and the aluminium alloy (73 GPa, 0.33)³¹.

Test bed. We created a custom test bed with five DOF; three DOF control the Cartesian movement of the probe (\vec{x} , \vec{y} , \vec{z}) using linear guide rails (Barch Motion) with a precision of 0.05 mm, and two DOF set the orientation of the sensor (yaw, roll) using Dynamixel MX-64AT and MX-28AT servo motors with a rotational precision of 0.09°, which results in a translational precision of 0.2 mm at

the tip of the sensor. The probe is fabricated from an aluminium alloy and is rigidly attached to the Cartesian gantry via an ATI Mini40 force/torque sensor with a force precision of 0.01/0.01/0.02 N ($F_x/F_y/F_z$). Insight is held at the desired orientation, and the indenter is used to contact it at the desired location.

Data. Measurements are collected using our automated test bed to probe Insight in different locations. To obtain a variety of normal and shear forces, the indenter moves to a specified location, touches the outer surface and deforms it increasingly by moving normal to the surface with fixed steps of 0.2 mm. For each indentation level, the indenter also moves sideways to apply shear forces (normal/shear movement ratio 2:1). After a pause of 2 s to allow transients to dissipate, we simultaneously record the contact location, the indenter contact force vector from the test bed's force sensor, and the camera image from inside Insight. When the measured total force exceeds 1.6 N, the data collection procedure at this specified location terminates and restarts at another location. The contact location and measured force vector are combined to create the true force distribution map using the method described in Supplementary Section A.3. Images from Insight are captured using a Raspberry Pi 4 Model B and are collected and combined using a standard laptop.

Challenges with analytical data processing. Classical force estimation pipelines compute the contact forces from the surface displacement using elastic theory³⁷. The displacement map can be acquired by analytically reconstructing the normals of the sensing surface or numerically deriving the relative movement of labelled markers from the raw image captured by the camera, as done in refs. ^{22,33}. However, large deformations violate linearity between displacement and force. In addition, the over-moulding in our design creates an inhomogeneous surface, where the stiffness matrix is difficult to model accurately. Shear forces are visible as small lateral deformations that highly depend on the distance to the stiff skeleton. Moreover, the reconstruction of surface normals requires evenly distributed light, without shadows or internal reflections²¹. Tracking markers^{23,33,52} rather than a surface does not solve the fundamental problems with displacement-focused approaches.

Machine learning. A ResNet⁴⁸ structure is used as our machine-learning model. The data for single contact includes a total of 187,358 samples at 3,800 randomly selected initial contact locations. The data set is split into training, validation and test subsets with a ratio of 3:1:1 according to the locations. The data for posture estimation from gravity contains 16,000 measurements and is split in the same way. We use four blocks of ResNet to estimate the contact position and amplitude directly (Fig. 3), two blocks to estimate the force distribution map (Fig. 4) and four blocks to estimate the sensor posture (Fig. 5). The machine-learning models are all trained with a batch size of 64 for 32 epochs, using Adam with a learning rate of 0.001 for mean squared loss minimization. Supplementary Section A.4 provides more details about the structure of the machine-learning models that we use. The performance of the models with less training data is studied in Supplementary Information A.6 (Supplementary Fig. 7).

Operating speed. The current version of Insight is not optimized for processing speed. Images are captured using a Raspberry Pi 4 with a Python script and transmitted to a host computer (with a GeForce RTX 2080Ti GPU) via Gigabit Ethernet. Images have a size of 1,640 × 1,232 and are effectively transferred at 11 fps and downsized to 410 × 308 using a Python script. The image processing with the deep network for the force-map prediction runs at 10 fps in real time. We see multiple ways to increase the operating speed, ranging from optimized code to hardware improvements (for example, using an Intel Neural Compute Stick) to choosing a faster deep network.

Comparison to state-of-the-art sensors. How does Insight compare with other vision-based haptic sensors? Table 1 lists its performance along with that of thirteen selected state-of-the-art sensors; we first give an overview and then compare the designs. One of the earliest vision-based sensors is GelSight²¹, which has a thin reflective coating on top of a transparent elastomer layer supported by a flat acrylic plate. Lighting parallel to the surface allows tiny deformations to be detected using photometric stereo techniques. Further developments of this approach increased its robustness (GelSlim²²), achieved curved sensing surfaces with one camera (GelTip²⁸) and with five cameras (OmniTact²⁷), and included markers to obtain shear force information²². A different technique based on tracking of small beads inside a transparent elastomer is used by GelForce³³ and the Sfrazza and D'Andrea sensor^{23,52} to estimate normal and shear force maps. ChromaForce (not listed in Table 1) uses subtractive colour mixing to extract similar data from deformable optical markers in a transparent layer⁵³. The TacTip²⁵ sensor family uses a hollow structure with a soft shell, and it detects deformations on the inside of that shell by visually tracking markers. Muscularis⁵⁴ and TacLink⁵⁵ extend this method to larger surfaces, such as robotic links, by using a pressurized chamber to maintain the shape of the outer shell; they are not listed in the table because they target a different application domain.

In terms of shape recognition and level of detail, the GelSight approach provides unparalleled performance. The tracking-based methods, such as GelForce

and TacTip, are naturally limited by their marker density and thicker outer layer. Insight uses shading effects to achieve a much higher information density than is possible with markers, but its accuracy is also somewhat limited by measuring at the inside of a soft shell with non-negligible thickness. Beyond accurately sensing contacts, the robustness of haptic sensors is of prime importance. Without additional protection, GelSight-based sensors are comparably fragile due to their thin reflective outer coating, which can easily be damaged. Adding another layer increases robustness, but imaging artefacts were reported to appear after about 1,500 contact trials due to wear effects⁴². We tested Insight for more than 400,000 interactions without noticeable damage or change in performance.

Each sensing technology imposes different restrictions on the surface geometry of the sensor. Vision-based tactile sensors need the measurement surface to be visible from the inside, so there is typically no space available for other items inside the sensor. The type of visual processing also matters. TacTip's need to track individual markers requires a more perpendicular view of the surface than shading-based approaches (GelSight and Insight). Soft materials deform well during gentle and moderate contact, but they do not withstand high forces if not adequately supported. GelSight uses a transparent rigid structure for support, which can lead to reflection artefacts when adapted to a curved sensing surface²⁸. An alternative is high internal pressure⁵⁵, but then the observed deformations are non-local. The over-moulded stiff skeleton in Insight maintains locality of deformations and withstands high forces.

To facilitate widespread adoption, tactile sensors need to be easy to produce from inexpensive components. Imaging components are remarkably cheap these days, making vision-based sensors competitive; however, GelSight needs a reproducible surface coating and permanent bonding between all layers, which are tricky to implement correctly^{21,56,57}. TacTip needs well-placed markers or a multi-material surface that can be three-dimensionally printed only by specialized machines. Insight uses one homogeneous elastomer that requires only a single-step moulding procedure on top of the stiff three-dimensionally printed skeleton. Being able to replace the sensing surface in a modular way increases system longevity; such replacement is supported by GelSight and TacTip in principle, and it is designed to be easy in Insight, although we did not evaluate the quality of the results that can be obtained without retraining.

Data availability

The data that support the findings of this study are available at <https://doi.org/10.17617/3.6c> (ref. ⁵⁸). The data comprise raw images and the corresponding contact information.

Code availability

Code used for training models and performing analyses are available at <https://doi.org/10.17617/3.6c> (ref. ⁵⁸).

Received: 14 July 2021; Accepted: 2 December 2021;

Published online: 23 February 2022

References

- Shah, K., Ballard, G., Schmidt, A. & Schwager, M. Multidrone aerial surveys of penguin colonies in Antarctica. *Sci. Robot.* **5**, abc3000 (2020).
- Nygård, T. F., Martin, C. P., Torresen, J., Glette, K. & Howard, D. Real-world embodied AI through a morphologically adaptive quadruped robot. *Nat. Mach. Intell.* **3**, 410–419 (2021).
- Ichnowski, J., Avigal, Y., Satish, V. & Goldberg, K. Deep learning can accelerate grasp-optimized motion planning. *Sci. Robot.* **5**, abd7710 (2020).
- Jain, S., Thiagarajan, B., Shi, Z., Clabaugh, C. & Mataric, M. J. Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Sci. Robot.* **5**, eaaz3791 (2020).
- Andrychowicz, M. et al. Learning dexterous in-hand manipulation. *Int. J. Robot. Res.* **39**, 3–20 (2020).
- Nagabandi, A., Konoglie, K., Levine, S. & Kumar, V. Deep dynamics models for learning dexterous manipulation. In *Proc. Conference on Robot Learning*, 1101–1112 (PMLR, 2020).
- Ballard, Z., Brown, C., Madni, A. M. & Ozcan, A. Machine learning and computation-enabled intelligent sensor design. *Nat. Mach. Intell.* **3**, 556–565 (2021).
- Yang, J. C. et al. Electronic skin: recent progress and future prospects for skin-attachable devices for health monitoring, robotics, and prosthetics. *Adv. Mater.* **31**, 1904765 (2019).
- Fishel, J. A. & Loeb, G. E. Sensing tactile microvibrations with the BioTac—comparison with human sensitivity. In *Proc. IEEE International Conference on Biomedical Robotics and Biomechatronics* 1122–1127 (IEEE, 2012).
- Lee, H., Park, H., Serhat, G., Sun, H. & Kuchenbecker, K. J. Calibrating a soft ERT-based tactile sensor with a multiphysics model and sim-to-real transfer learning. In *Proc. IEEE International Conference on Robotics and Automation* 1632–1638 (IEEE, 2020).
- Chen, M. et al. An ultrahigh resolution pressure sensor based on percolative metal nanoparticle arrays. *Nat. Commun.* **10**, 4024 (2019).

12. Sun, H. & Martius, G. Machine learning for haptics: inferring multi-contact stimulation from sparse sensor configuration. *Front. Neurorobot.* **13**, 51 (2019).
13. Taunyazov, T. et al. Event-driven visual-tactile sensing and learning for robots. In *Proc. Robotics: Science and Systems* (2020).
14. Boutry, C. M. et al. A hierarchically patterned, bioinspired e-skin able to detect the direction of applied pressure for robotics. *Sci. Robot.* **3**, aau6914 (2018).
15. Mittendorfer, P. & Cheng, G. Humanoid multimodal tactile-sensing modules. *IEEE Trans. Robot.* **27**, 401–410 (2011).
16. Guadarrama-Olvera, J. R., Bergner, F., Dean, E. & Cheng, G. Enhancing biped locomotion on unknown terrain using tactile feedback. In *Proc. IEEE International Conference on Humanoid Robots (Humanoids)* 1–9 (IEEE, 2018).
17. Park, J., Kim, M., Lee, Y., Lee, H. S. & Ko, H. Fingertip skin-inspired microstructured ferroelectric skins discriminate static/dynamic pressure and temperature stimuli. *Sci. Adv.* **1**, 1500661 (2015).
18. Lai, Y.-C., Hsiao, Y.-C., Wu, H.-M. & Wang, Z. L. Waterproof fabric-based multifunctional triboelectric nanogenerator for universally harvesting energy from raindrops, wind, and human motions and as self-powered sensors. *Adv. Sci.* **6**, 1801883 (2019).
19. Piacenza, P., Behrman, K., Schifferer, B., Kymmissis, I. & Ciocarlie, M. A Sensorized multicurved robot finger with data-driven touch sensing via overlapping light signals. *IEEE/ASME Trans. Mechatronics* **25**, 2416–2427 (2020).
20. Bai, H. et al. Stretchable distributed fiber-optic sensors. *Science* **370**, 848–852 (2020).
21. Yuan, W., Dong, S. & Adelson, E. H. GelSight: high-resolution robot tactile sensors for estimating geometry and force. *Sensors* **17**, 2762 (2017).
22. Ma, D., Donlon, E., Dong, S. & Rodriguez, A. Dense tactile force estimation using GelSlim and inverse FEM. In *Proc. IEEE International Conference on Robotics and Automation* 5418–5424 (IEEE, 2019).
23. Sferrazza, C. & D'Andrea, R. Design, motivation and evaluation of a full-resolution optical tactile sensor. *Sensors* **19**, 928 (2019).
24. Van Duong, L., Asahina, R., Wang, J. & Ho, V. A. Development of a vision-based soft tactile muscularis. In *Proc. IEEE International Conference on Soft Robotics* 343–348 (IEEE, 2019).
25. Ward-Cherrier, B. et al. The TacTip family: soft optical tactile sensors with 3D-printed biomimetic morphologies. *Soft Robot.* **5**, 216–227 (2018).
26. Lee, B. et al. Ultraflexible and transparent electroluminescent skin for real-time and super-resolution imaging of pressure distribution. *Nat. Commun.* **11**, 663 (2020).
27. Padmanabha, A. et al. OmniTact: a multi-directional high-resolution touch sensor. In *Proc. IEEE International Conference on Robotics and Automation* 618–624 (IEEE, 2020).
28. Gomes, D. F., Lin, Z. & Luo, S. GelTip: a finger-shaped optical tactile sensor for robotic manipulation. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems* 9903–9909 (IEEE, 2020).
29. Romero, B., Veiga, F. & Adelson, E. Soft, round, high resolution tactile fingertip sensors for dexterous robotic manipulation. In *2020 IEEE International Conference on Robotics and Automation* 4796–4802 (IEEE, 2020).
30. Lee, H., Chung, J., Chang, S. & Yoon, E. Normal and shear force measurement using a flexible polymer tactile sensor with embedded multiple capacitors. *J. Microelectromechanical Syst.* **17**, 934–942 (2008).
31. Yan, Y. et al. Soft magnetic skin for super-resolution tactile sensing with force self-decoupling. *Sci. Robot.* **6**, abc8801 (2021).
32. Lambeta, M. et al. DIGIT: a novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robot. Automation Lett.* **5**, 3838–3845 (2020).
33. Sato, K., Kamiyama, K., Nii, H., Kawakami, N. & Tachi, S. Measurement of force vector field of robotic finger using vision-based haptic sensor. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems* 488–493 (IEEE, 2008).
34. Tres, P. A. in *Designing Plastic Parts for Assembly* 327–341 (Hanser, 2014).
35. Woodham, R. J. Photometric method for determining surface orientation from multiple images. *Opt. Eng.* **19**, 139–144 (1980).
36. Geng, J. Structured-light 3D surface imaging: a tutorial. *Adv. Opt. Photon.* **3**, 128–160 (2011).
37. Perez, N. *Theory of Elasticity* 1–52 (Springer, 2017).
38. Deimel, R. & Brock, O. A novel type of compliant and underactuated robotic hand for dexterous grasping. *Int. J. Robot. Res.* **35**, 161–185 (2016).
39. Schmitt, F., Piccin, O., Barbé, L. & Bayle, B. Soft robots manufacturing: a review. *Front. Robot. AI* **5**, 84 (2018).
40. Albu-Schaffer, A. et al. Soft robotics. *IEEE Robotics Automation Magazine* **15**, 20–30 (2008).
41. Rus, D. & Tolley, M. T. Design, fabrication and control of soft robots. *Nature* **521**, 467–475 (2015).
42. Donlon, E. et al. GelSlim: a high-resolution, compact, robust, and calibrated tactile-sensing finger. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems* 1927–1934 (IEEE, 2018).
43. Shahnewaz, A. & Pandey, A. K. *Color and Depth Sensing Sensor Technologies for Robotics and Machine Vision* 59–86 (Springer, 2020).
44. Aliaga, D. G. & Xu, Y. Photogeometric structured light: a self-calibrating and multi-viewpoint framework for accurate 3D modeling. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 1–8 (IEEE, 2008).
45. Shao, Y., Hayward, V. & Visell, Y. Compression of dynamic tactile information in the human hand. *Sci. Adv.* **6**, aaz1158 (2020).
46. Van Duong, L., Asahina, R., Wang, J. & Ho, V. A. Development of a vision-based soft tactile muscularis. In *Proc. IEEE International Conference on Soft Robotics* 343–348 (IEEE, 2019).
47. Li, Q. et al. A review of tactile information: perception and action through touch. *IEEE Trans. Robot.* **36**, 1–16 (2020).
48. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
49. Kuhn, H. W. & Yaw, B. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* <https://doi.org/10.1002/nav.3800020109> (1955).
50. Academic Research Mechanical Release 18.1 (Ansys, 2020).
51. Mott, P. H. & Roland, C. M. Limits to Poisson's ratio in isotropic materials—general result for arbitrary deformation. *Physica Scripta* **87**, 055404 (2013).
52. Sferrazza, C., Bi, T. & D'Andrea, R. Learning the sense of touch in simulation: a sim-to-real strategy for vision-based tactile sensing. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems* 4389–4396 (IEEE, 2020).
53. Lin, X. & Wiertlewski, M. Sensing the frictional state of a robotic skin via subtractive color mixing. *IEEE Robot. Automation Lett.* **4**, 2386–2392 (2019).
54. Van Duong, L., Asahina, R., Wang, J. & Ho, V. A. Development of a vision-based soft tactile muscularis. In *Proc. IEEE International Conference on Soft Robotics* 343–348 (IEEE, 2019).
55. Van Duong, L. & Ho, V. A. Large-scale vision-based tactile sensing for robot links: design, modeling, and evaluation. *IEEE Trans. Robot.* **37**, 1–14 (2020).
56. Dong, S., Yuan, W. & Adelson, E. H. Improved GelSight tactile sensor for measuring geometry and slip. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems* 137–144 (IEEE, 2017).
57. Wilson, A., Wang, S., Romero, B. & Adelson, E. *Design of a Fully Actuated Robotic Hand With Multiple Gelsight Tactile Sensors* (2020).
58. Sun, H., Kuchenbecker, K. J. & Martius, G. *Data & Code for Insight: A Haptic Sensor Powered by Vision and Machine Learning* (2021); <https://doi.org/10.17617/3.6c>
59. Narang, Y., Wyk, K. V., Mousavian, A. & Fox, D. Interpreting and predicting tactile signals via a physics-based and data-driven framework. In *Proc. Robotics: Science and Systems* (2020).
60. Molchanov, A., Kroemer, O., Su, Z. & Sukhatme, G. S. Contact localization on grasped objects using tactile sensing. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 216–222 (IEEE, 2016).
61. Johnson, K. L. *Normal Contact of Elastic Solids—Hertz Theory* 84–106 (Cambridge University Press, 1985).

Acknowledgements

We thank the China Scholarship Council (CSC) and the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting H.S. G.M. is member of the Machine Learning Cluster of Excellence, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC no. 2064/1) project no. 390727645. We acknowledge the support from the German Federal Ministry of Education and Research (BMBF) through the Tübingen AI Center (FKZ: 01IS18039B).

Author contributions

H.S., K.J.K. and G.M. conceived the method and the experiments, drafted the manuscript and revised it. H.S. designed and constructed the hardware, developed fabrication methods, designed and conducted experiments, collected and analysed the data. G.M. and K.J.K. supervised the data analysis. We thank F. Grimminger and B. Javot for their support during the mechanical manufacturing procedure.

Funding

Open access funding provided by Max Planck Society

Competing interests

H.S., G.M. and K.J.K. are listed as inventors on two PCT provisional patent applications (PCT/EP2021/050230, PCT/EP2021/050231 filed internationally (PCT)) that cover the fundamental principles and designs of Insight.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s42256-021-00439-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42256-021-00439-3>.

Correspondence and requests for materials should be addressed to Huabo Sun or Georg Martius.

Peer review information *Nature Machine Intelligence* thanks Daniel Fernandes Gomes and Benjamin Ward-Cherrier for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

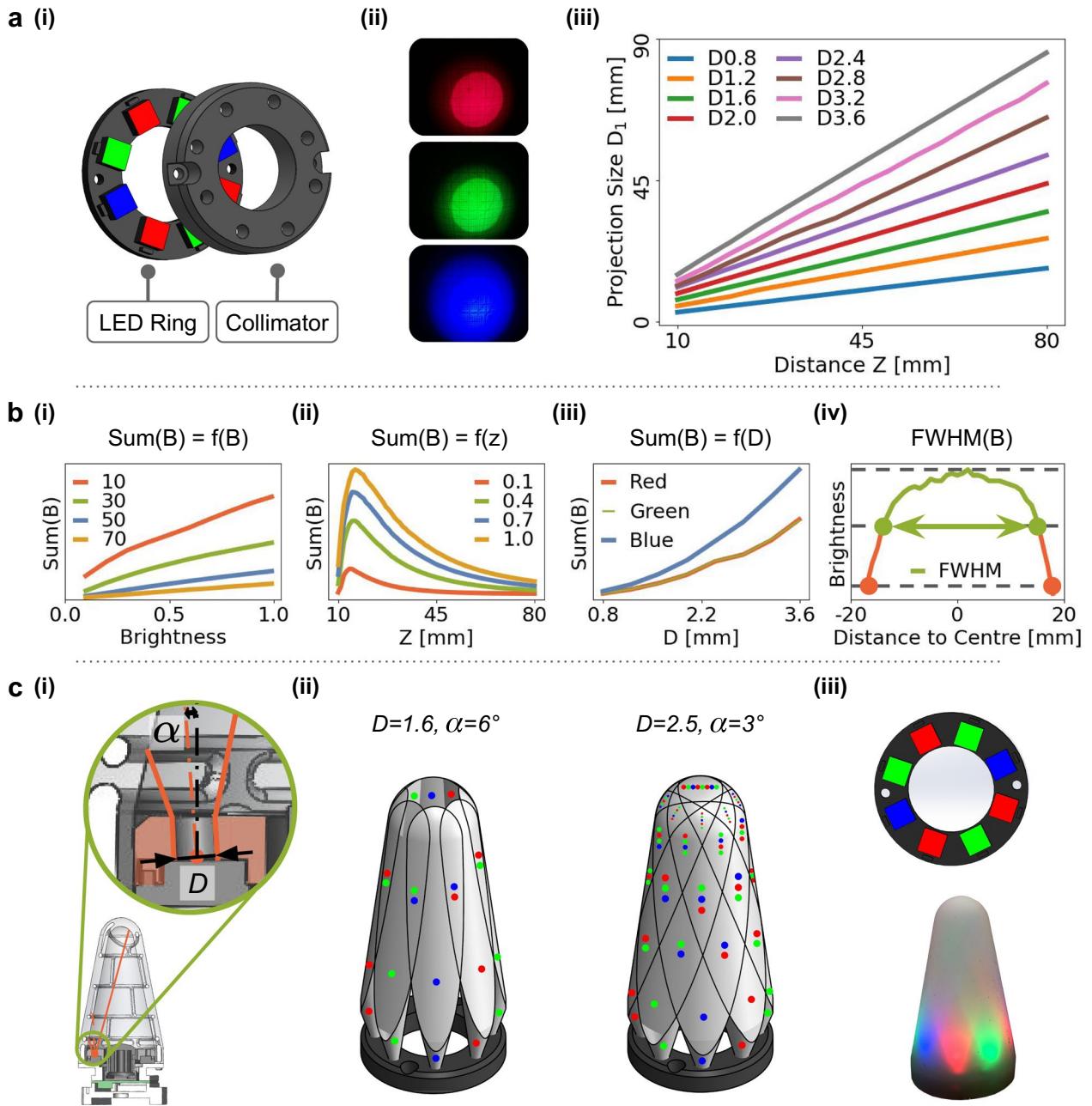
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



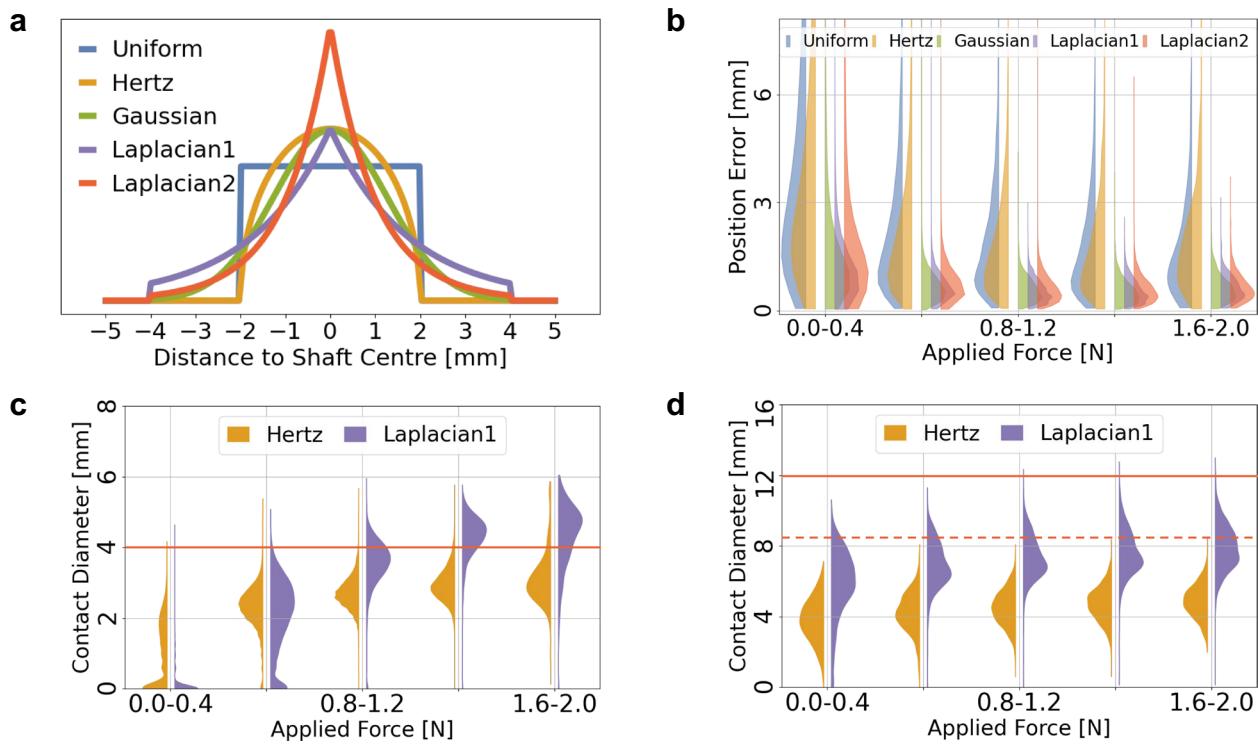
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

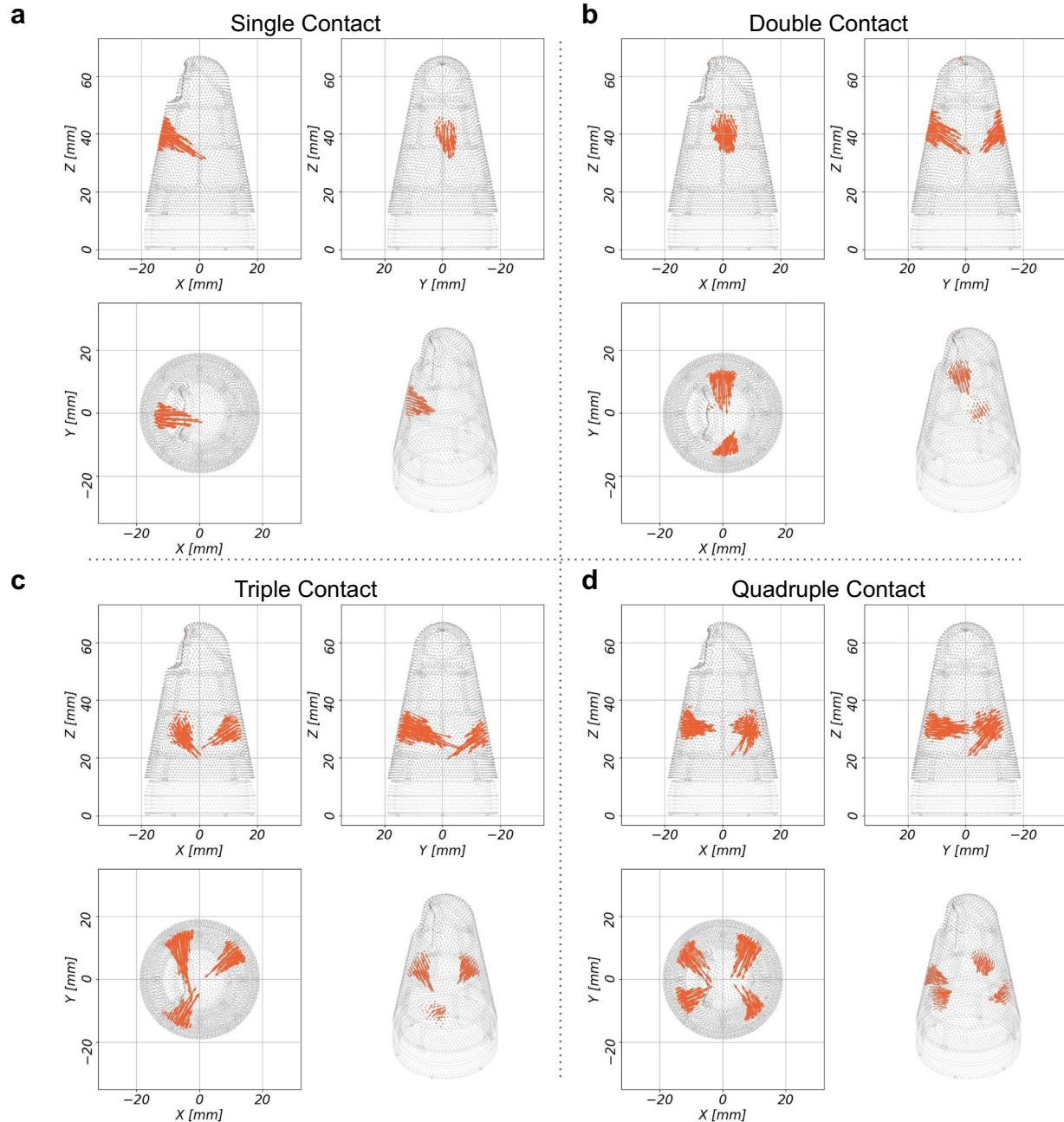
© The Author(s) 2022



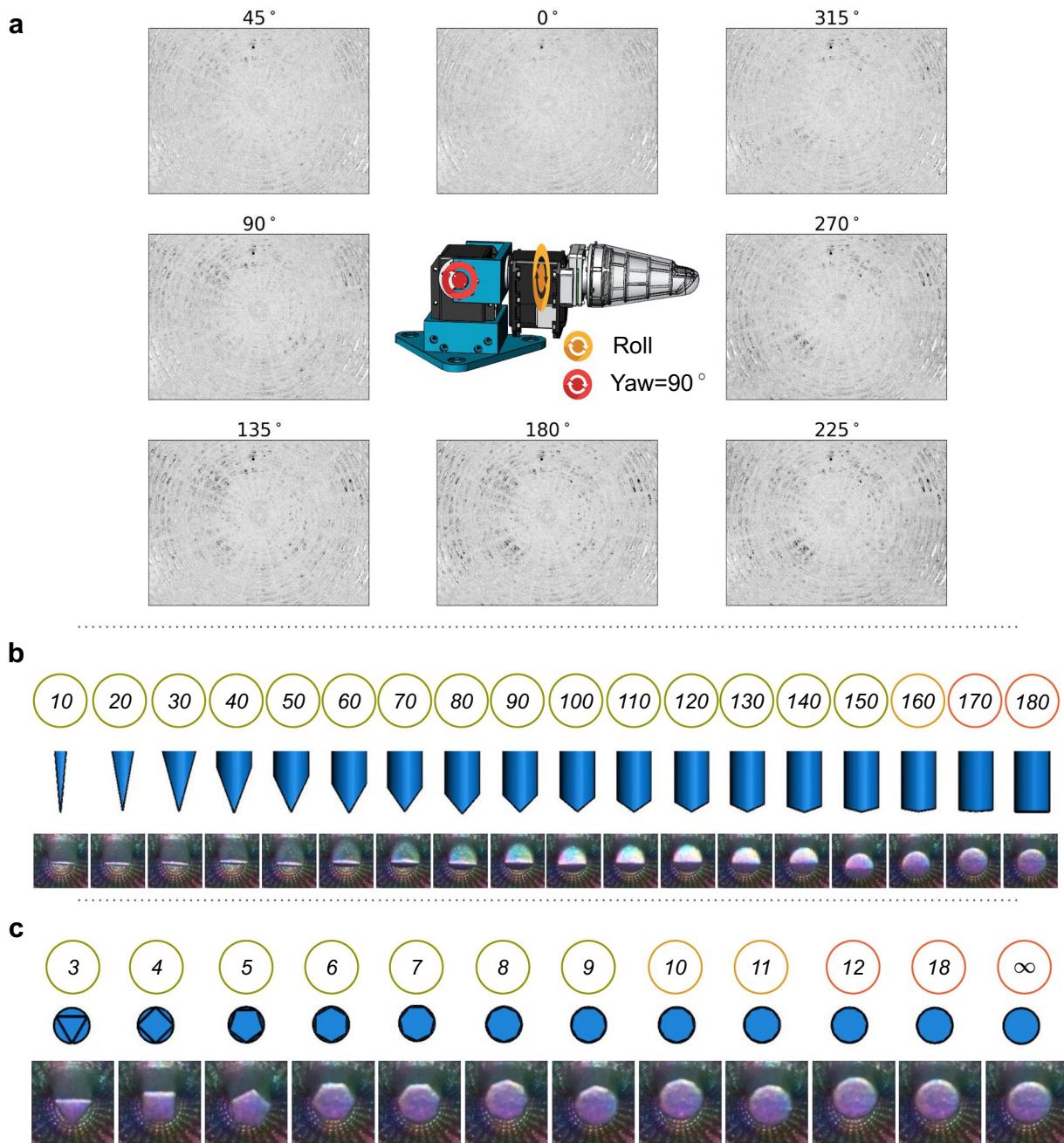
Extended Data Fig. 1 | Analysis of the lighting system. **a**, The correlation between collimator diameter D and light cone size. **a(i,ii)**, The LED ring, the collimator, and light patterns projected on a flat plane from different red, green, and blue light sources. **a(iii)**, The linear scaling of the cone size. Z is the imaging distance, D_1 is the diameter of the projection pattern, and different lines (D) are for different collimator diameters. **b**, A summary of the light attenuation behaviour. **b(i)**, The sum of light as seen by the camera depends on the surface distances (Z) for different brightness values of a single light source. **b(ii)**, The same quantity as a function of the imaging distance for different brightness values. **b(iii)**, The camera sensitivity to light brightness of different colours (red, green, blue) with varying D . **b(iv)**, The light attenuation curve for a single bright disc (as in **a(ii)**). We use full width at half maximum (FWHM) to quantify the size of the disc. **c**, The effect of collimator hole size D and angle α on the light cones and the overall light pattern. **c(i)**, The details of the collimator hole geometry. **c(ii)**, The light covers the surface area with different D and α . **c(iii)**, The light colour arrangement (R, G, B, R, G, R, B, G) and visibility in a translucent shell.



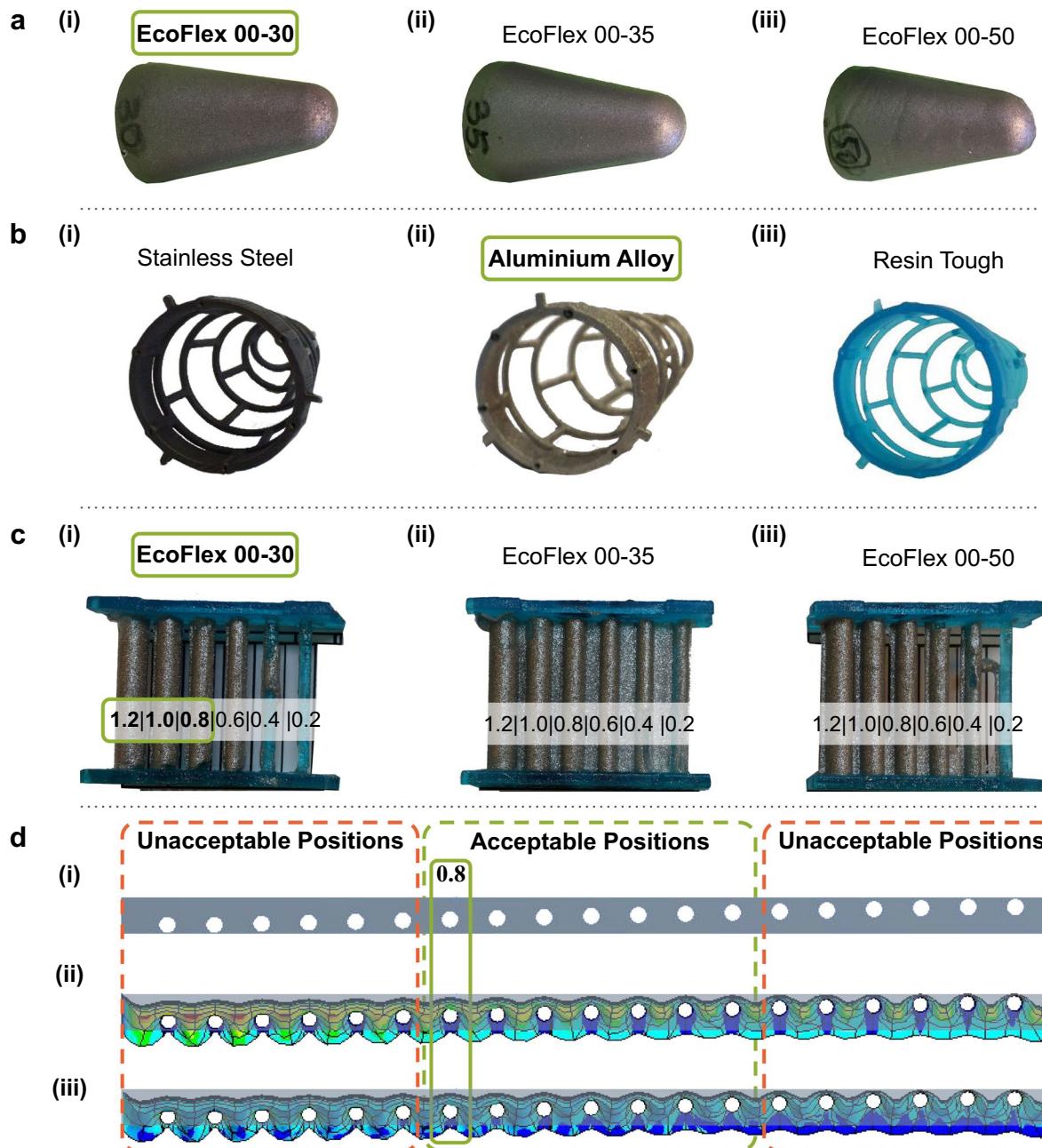
Extended Data Fig. 2 | Force distribution approximation and evaluation. **a**, The five approximate force distribution methods that we tested for the 4 mm sphere-shaped indenter. The illustration is in one dimension and was revolved to distribute the measured contact force vector across the local surface of the force map around the contact point. The uniform (blue) and Hertz theory⁶¹ (orange) distribution curves are strongly localized with a radius of 2 mm. The green curve follows a truncated Gaussian distribution, and the two other curves (purple and red) follow Laplacian distributions that also stop at a radius of 4 mm. **b**, The position inference performance using the different methods. **c** and **d**, The diameters of the contact area prediction for two indenter sizes (4 mm and 12 mm) using two different approximation methods (Hertz and Laplacian1). The dashed line indicates a reference diameter from Supplementary Fig. 5 **ai** based on the fact that the large indenter penetrates into the sensor only partially.



Extended Data Fig. 3 | Examples of multiple contacts. Visualizations of the force map distributions over the sensing surface for a single contact (a), double contact (b), triple contact (c), and quadruple contact (d).



Extended Data Fig. 4 | Sensitivity evaluation. **a**, The image changes caused by gravity when the sensor rotates 360° around the roll direction while maintaining a yaw angle of 90°. In contrast to the images actually used for the posture detection experiment, the images presented here were recorded in typical overhead lighting conditions. Nevertheless we see no illumination impact even at the thin fovea part, showing the skin is sufficiently opaque. **b** and **c** extend the reported evaluation of the sensitivity of shape detection for wedge sharpness and polygon edges.



Extended Data Fig. 5 | Mechanical aspects of the sensor design. **a**, The candidate materials for the soft sensing shell: we choose EcoFlex 00-30. **b**, The candidate materials for the stiff skeleton: we choose aluminium alloy. **c**, The thickness test for the over-moulding technique, which is used to connect the elastomer and the skeleton without adhesive. The minimal thickness for a robust connection is 0.8 mm. **d**, A finite element analysis on how the relative position of the skeleton in the elastomer will affect the sensitivity, i.e. how much deformation occurs. **di**, A soft plate containing stiff rods at varying distances to the upper and lower surfaces. The left and right edges and the rods are fixed. Homogeneous pressure is applied to the upper surface. **di**, The resulting deformation. **di**, The induced von Mises stress.