nature sensors

Article

https://doi.org/10.1038/s44460-025-00006-y

Biomimetic multimodal tactile sensing enables human-like robotic perception

Received: 25 April 2025

Accepted: 12 November 2025

Published online: 15 January 2026

Check for updates

Shoujie Li[1,15], Tong Wu[1,15], Jianle Xu[1,15], Yan Huang[1], Zongwen Zhang[2,3], Hongfa Zhao[1], Qinghao Xu[1], Zihan Wang[4], Linqi Ye[5], Yang Yang[6], Chuqiao Lyu[1], Shilong Mu[1,7], Xueqian Wang[1], Zhaoqian Xie[2,3] ✉, Changsheng Wu[8,9,10,11] ✉, Xinge Yu[12,13,14] ✉ & Wenbo Ding[1] ✉

Tactile perception is essential for skilled robotic manipulation, yet current systems are limited by low sensor resolution, incomplete modality integration and insufficient interpretation of complex tactile signals. Here we present the Superior Tactile Sensor (SuperTac), a biomimetic, multimodal tactile sensor inspired by the multispectral vision of pigeons. SuperTac integrates multispectral imaging (MIR to ultraviolet light) with triboelectric and inertial sensing into a single 1-mm-thick light-field-modulated skin comprising conductive polymer, fluorescent, reflective and supporting layers. The sensor combines pressure-adaptive force sensing with high-resolution (0.00545 mm^2 px^{-1}) and high-precision measurements across force (0.06 N accuracy), position (0.4 mm accuracy), temperature (0–90 °C range), proximity (<15 cm range) and vibration (0–60 Hz range). It achieves over 94% accuracy in discriminating texture, material, sliding, collision and colour. To interpret these rich multimodal data, we developed DOVE, an 8.5B-parameter tactile language model that enables sophisticated understanding of tactile stimuli. This integrated sensing and interpretation framework could bring robotic touch perception closer to human-like capabilities, with potential applications in manufacturing, healthcare and service robotics.

Touch is a fundamental sensory modality for robotic manipulation[1], human–robot interaction (HRI)[2] and extended reality[3] applications. As embodied intelligence has advanced, the demand for sophisticated tactile sensing capabilities has grown exponentially. High-resolution multimodal tactile sensors, capable of detecting fine object details while capturing diverse physical information, have emerged as a critical focus in both academic research and industrial development[4,5].

Tactile sensors based on electronic skin (e-skin) initially demonstrated notable potential for multimodal sensing due to their versatile functional materials[6–8]. However, increasing spatial resolution and sensing modalities in e-skin necessitates denser electrode arrays, resulting in signal crosstalk and complex readout circuitry. In contrast, visuotactile sensing has been proposed as an elegant alternative, offering sub-millimetre spatial resolution through optical imaging while naturally integrating with modern artificial intelligence frameworks, including computer vision[9], deep neural networks[4] and large language models[10,11]. Despite these advantages, extending visuotactile sensing to incorporate multispectral and non-imaging modalities presents substantial technical challenges. Although traditional visual systems can readily integrate non-visible light sensors, this approach is hindered in visuotactile systems due to constraints imposed by the sensing skin. Although recent advancements have demonstrated bimodal visuotactile sensors capable of simultaneous temperature-force[4] and material-force[12] sensing, most implementations remain confined to the visible spectrum (Supplementary Tables 1 and 2). Consequently, the development of truly multimodal visuotactile sensors

A full list of affiliations appears at the end of the paper. ✉e-mail: zxie@dlut.edu.cn; cwu@nus.edu.sg; xingeyu@cityu.edu.hk; ding.wenbo@sz.tsinghua.edu.cn

https://doi.org/10.1038/s44460-025-00006-y

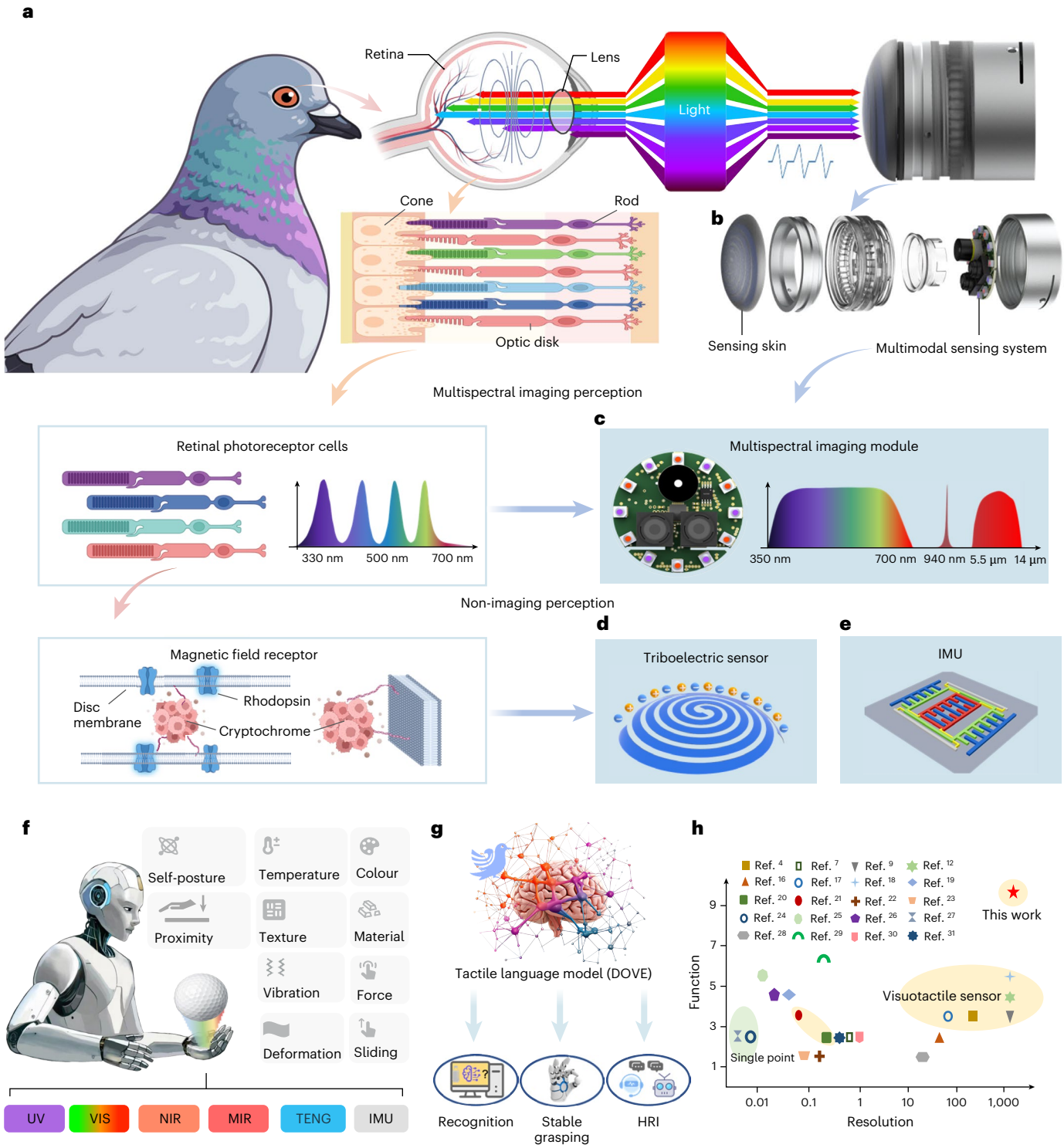

Fig. 1 | Overview of the multimodal tactile sensing system. a, The structure of the retina in pigeons includes cones and rods. We drew inspiration from their remarkable multispectral vision, along with specialized retinal molecules for non-imaging perception, such as magnetic field detection. **b**, The overall structure of the sensor comprises a sensing skin and a multimodal sensing system. **c**, Multispectral imaging systems achieve visible, ultraviolet, NIR and MIR spectral sensing. **d**,**e**, A triboelectric sensor (**d**) and an IMU (**e**) enhance the sensing capability of the tactile sensor. **f**, SuperTac's demonstration of sensing modalities and functions. Deploying sensors with a manipulator can enable the sensing of ten functions. **g**, SuperTac, combined with the tactile language model (DOVE), can be applied in object recognition, grasping and HRI. **h**, Comparison of the resolution and functionality of current mainstream tactile sensors (data from refs. 4,7,9,12,16–31). Green shading indicates that the literature only designed a single sensing module without an array structure. Yellow shading indicates that a visuotactile sensing method was adopted. TENG, triboelectric nanogenerator; UV, ultraviolet; VIS, visible. Panel **a** created with BioRender.com.

faces two primary obstacles: limitations in sensing skin design and a restricted imaging bandwidth.

Inspired by the remarkable multispectral vision of pigeons[13,14], we introduce the Superior Tactile Sensor (SuperTac) (Fig. 1a,b and Supplementary Videos 1 and 2), an integrated multimodal high-resolution (0.00545 mm^2 px^{-1}) tactile sensor that combines multispectral imaging (Fig. 1c), triboelectric sensing (Fig. 1d) and inertial measurement (Fig. 1e). At the heart of SuperTac is a miniaturized sensing unit

(Supplementary Note 1 and Supplementary Table 3) featuring light-field modulation and multispectral imaging capabilities. The sensor employs a transparency-adjustable multilayered sensing skin comprising a poly(3,4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT:PSS)[15] conductive layer, an ultraviolet ink fluorescent layer and a silver-powder-coated reflective layer. This design enables different functional modes across various spectra through light-field modulation. Additionally, an integrated inertial measurement unit (IMU) provides complementary acceleration and posture data. SuperTac achieves comprehensive sensing capabilities, including force, texture, deformation, temperature, sliding, material properties, distance, vibrations, collision detection and colour recognition (Fig. 1f and Supplementary Video 3). A unique feature of the sensor is its adjustable internal air pressure, which allows for dynamic adaptation of the force sensing range. Through deep learning integration, SuperTac shows exceptional performance: a force measurement accuracy of 0.06 N, a position accuracy of 0.4 mm, a temperature range of 0–90 °C, proximity detection, vibration sensing from 0–60 Hz, and over 94% accuracy in texture, material, sliding, collision and colour classification. To showcase its practical applications, we integrated SuperTac into a dexterous robotic hand and developed DOVE, a specialized tactile language model. DOVE accurately interprets tactile information from manipulated objects, indicating the sensor's potential for advanced HRI and robotic manipulation tasks (Fig. 1g). This integrated approach achieves excellent resolution and functionality compared with existing solutions[4,7,9,12,16–31] (Fig. 1h).

Bio-inspired design of the multimodal tactile sensor

The vertebrate retina contains specialized photoreceptors—rods and cones—with cones enabling colour vision. Unlike humans, pigeons possess an additional type of cone cell that is sensitive to ultraviolet wavelengths[32], along with specialized retinal molecules for non-imaging perception, such as magnetic field detection[33]. This enhanced visual system enables pigeons to process complex environmental information more comprehensively. Drawing inspiration from these capabilities, SuperTac combines multispectral imaging with triboelectric and inertial sensing to expand the perceptual capabilities of visuotactile sensors. Based on this design, through a single touch, the sensor can obtain information about an object's shape, texture, colour, temperature and material, as well as the force during contact.

Structural design and sensing mechanism

Visuotactile sensing, which utilizes vision for tactile perception[34], has become increasingly valuable for robotic grasping[35] and manipulation[36], particularly given its compatibility with foundation model frameworks, such as the vision–language–action model[37]. Traditional visuotactile sensors typically comprise sensing skin, imaging and lighting modules. In contrast, SuperTac introduces an innovative design that integrates multispectral imaging, triboelectric signal acquisition, IMU signal acquisition and lighting modules into a unified multimodal sensing system, greatly enhancing both functionality and integration. This integrated design enables comprehensive environmental interaction through multiple sensing modalities (Fig. 2a). The system can simultaneously detect force, texture, deformation, temperature, material properties, proximity, sliding, pose, vibration and colour (Supplementary Tables 1 and 2), providing a detailed multisensory representation of physical interactions.

The sensor's design combines multiple functional elements (Fig. 2b). The core innovative part is a multilayer transparent sensing skin coupled with a multimodal sensing system capable of precise spectral band detection, triboelectric signal acquisition and IMU-based motion sensing. To capture triboelectric signals, we developed a transparent conductive layer based on PEDOT:PSS integrated into the sensing skin. The design also incorporates an IMU for orientation and acceleration sensing. These components are compactly integrated into a four-layer printed circuit board with a radius of 16 mm, housing the multispectral imaging, triboelectric, IMU signal acquisition and lighting modules (Supplementary Note 2 and Supplementary Figs. 1–3).

Sensing skin

The selection and structure of sensing skin materials are optimized to enhance SuperTac's functionalities (Supplementary Note 3). The skin comprises four layers: a conductive layer, a reflective layer, a fluorescent layer and a supporting layer (Fig. 2b and Supplementary Fig. 4), with a thickness of only 1 mm (Supplementary Fig. 5). The conductive layer, fabricated by screen printing transparent PEDOT:PSS ink on thermoplastic polyurethane (TPU) thin film, generates triboelectric signals during object contact. PEDOT:PSS provides excellent transparency and conductivity, whereas TPU offers exceptional stretchability, transparency and toughness (Supplementary Fig. 6). The combination ensures both film transparency and stable triboelectric signal generation. The electrode adopts a vortex line (PEDOT:PSS) design to provide a uniform signal. Based on the triboelectric mechanism (Supplementary Note 4), the conductive layer generates distinct electrical signals upon contact with objects of varying electronegativities, enabling material-type discrimination and proximity sensing (Supplementary Fig. 7).

The reflective layer operates similarly to a one-way mirror (Fig. 2c and Supplementary Figs. 8 and 9). Its transparency is regulated by light intensity on either side: on the bright side, reflected light dominates, rendering the film opaque, whereas on the dark side, transmitted light prevails, making the film transparent. This mechanism allows independent imaging across different wavelengths by controlling the light intensity in specific spectral bands.

The fluorescent layer employs ultraviolet light to control marker visibility. These markers, visible in the ultraviolet spectrum but invisible in the near-infrared (NIR) band, enable the sensor to alternate between detection modes with and without markers (Supplementary Fig. 4). This capability allows simultaneous deformation and slide detection without compromising texture detection. When combined with the multispectral imaging system, it captures ultraviolet markers and NIR texture information.

The supporting layer is the base substrate of the sensing skin, providing mechanical integrity and structural stability for the entire multilayer assembly. Its main functions are to maintain the overall shape and flexibility of the skin, ensure reliable integration and alignment of the other functional layers (conductive, reflective and fluorescent) and protect the sensor from mechanical damage during repeated deformations. Additionally, the supporting layer serves as a physical barrier, isolating the functional layers from external contaminants and environmental factors, thereby enhancing the durability and longevity of the sensor. Unlike traditional acrylic-based designs, we employ a silicone-based inflatable support structure. This design offers several advantages: a larger deformation range for detailed object contour representation; a force sensing range (0–7 N) that is adjustable through internal air pressure control (Supplementary Fig. 10); and an improved thermal response due to its thinner profile. Additionally, the silicone inflatable film addresses the limitations of mid-infrared (MIR; 5.5–14.0 μm wavelength) temperature sensing, where traditional materials such as acrylic and standard glass cannot transmit wavelengths above 5 μm. This eliminates the need for costly, special optical glass while maintaining performance. However, the pneumatic support structure offers advantages such as adjustable pressure sensing and enhanced deformation sensing but poses challenges related to sealing, material ageing and repeatability. To address these issues, we integrated a compact air supply system, replaced latex with durable silicone and utilized TPU film for improved wear resistance, achieving superior durability and consistent performance over 80,000 tests.

Multimodal sensing system

The multimodal sensing system integrates four modules: multispectral imaging, triboelectric signal acquisition, IMU signal acquisition and

https://doi.org/10.1038/s44460-025-00006-y

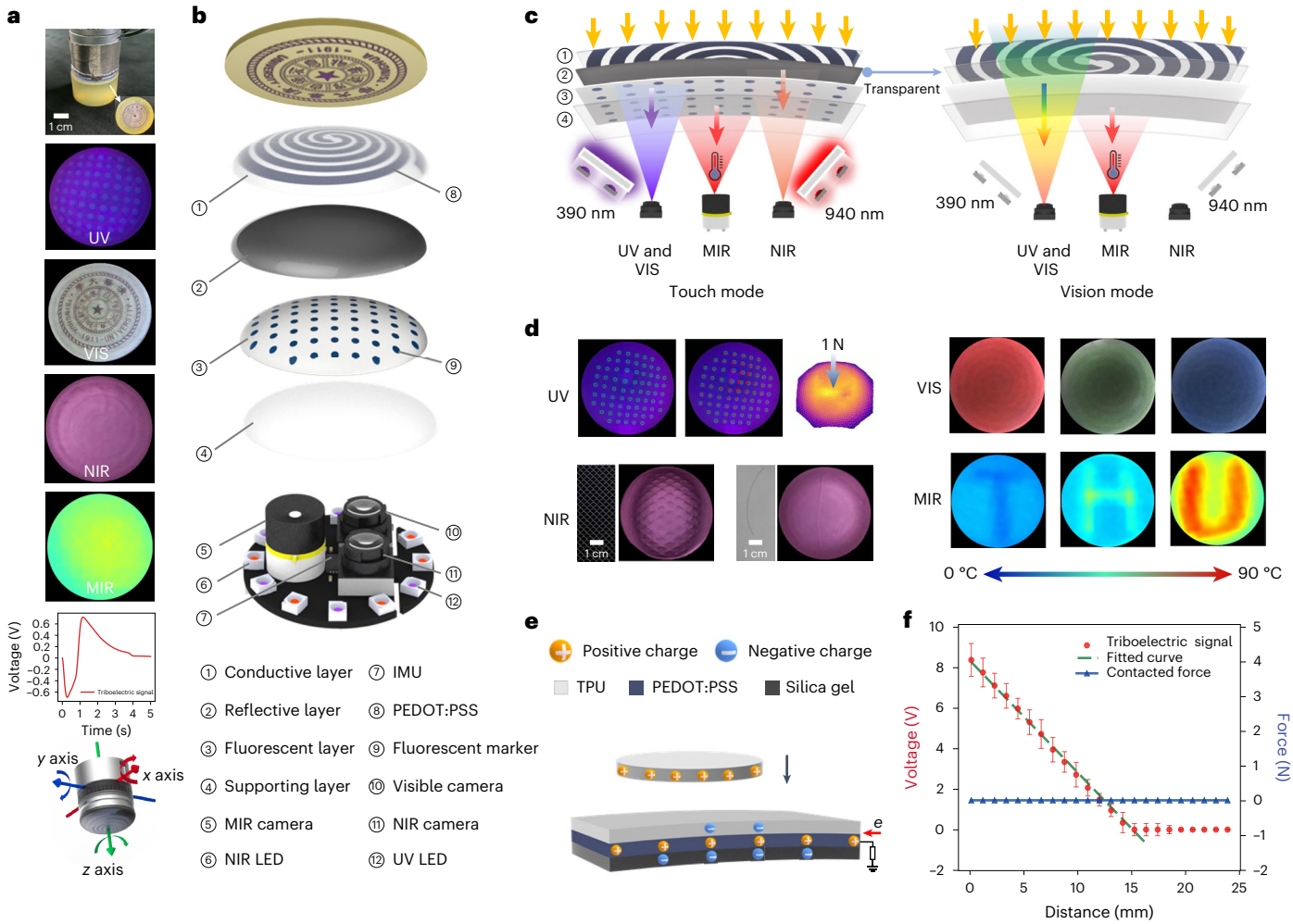

Fig. 2 | Structural design and sensing mechanism. a, Sensing modalities of SuperTac. **b**, Structure of the sensing skin, imaging module and lighting module. **c**, SuperTac is in touch mode when the internal lighting module is turned on and in vision mode when the internal lighting module is turned off. **d**, Tactile data are captured when the sensor is in contact with an object. Shown are representative ultraviolet images in touch mode (top left), NIR images in touch mode (bottom left), RGB images in vision mode (top right) and temperature data (bottom right). **e**, Triboelectric signal acquisition mechanism. **f**, Object proximity sensing. The error bars represent the maximum and minimum values of the error ($n = 5$ independent experimental repetitions per data point).

lighting modules (Fig. 2b). The miniaturized multispectral imaging module includes an MIR camera and two cameras incorporating a complementary metal-oxide semiconductor (CMOS)—one with low-pass filtering and the other with bandpass filtering. The system covers four spectral bands: ultraviolet (390 nm illumination and 450 nm fluorescence), visible (400–700 nm), NIR (940 nm) and MIR (5.5–14.0 μm) (Supplementary Fig. 11). To prevent crosstalk, the tactile mode uses ultraviolet fluorescence detection, whereas the visual mode captures external visible light with the ultraviolet light-emitting diode (LED) turned off (Fig. 2d).

MIR detection

For temperature measurement, we employ an MIR imaging camera (MLX90640) with 24 × 32 resolution that is capable of detecting wavelengths between 5.5 and 14.0 μm and measuring temperatures from −40 to 300 °C. This camera captures thermal radiation emitted by objects, enabling precise temperature mapping.

NIR detection

A CMOS unit paired with a 935- to 945-nm bandpass filter and lens provides precise NIR signal detection, with filter selection determined by the wavelength of the LED light source.

Visible and ultraviolet light detection

A CMOS unit with a 700 nm low-pass filter and lens covers an imaging range from 350–1,000 nm, encompassing ultraviolet, visible and NIR spectra. LED lighting adjustment enables selective wavelength detection.

The lighting module is meticulously designed to support both reflective and fluorescent layer functionalities. For fluorescent marker detection, 390 nm LEDs excite the fluorescent layer, revealing marker information. The ultraviolet fluorescent markers enable modality switching for deformation, sliding and texture sensing, offering advantages in three-dimensional (3D) reconstruction and sliding detection without relying on strict light control. When deactivated, the fluorescent layer becomes transparent, allowing external colour observation (Fig. 2c). For texture sensing, 940 nm LEDs generate a strong internal NIR light source, rendering the thin film opaque and enhancing surface texture detection (Supplementary Fig. 12). This light source also works in conjunction with the NIR detection unit, providing stable illumination for precise signal detection (Supplementary Note 5).

For triboelectric signal acquisition, we use an ADA4505 chip operating at a 1 kHz sampling frequency (Supplementary Table 4). The IMU signal acquisition utilizes MPU6050, capturing 3D orientation angles and acceleration data. This configuration enables comprehensive

multimodal sensing while maintaining system compactness and integration, addressing the limitations of traditional visuotactile sensors.

SuperTac demonstrates comprehensive sensing capabilities across multiple spectral bands and sensing modalities (Fig. 2d). In the ultraviolet band, fluorescent markers enable precise tracking of sliding and deformation through marker size and displacement measurements (Supplementary Note 6, Supplementary Table 5 and Supplementary Figs. 13–15). The visible spectrum provides object colour information upon contact, whereas the NIR band captures texture and contact-force data. MIR imaging enables temperature measurement, complemented by triboelectric signals for material identification (Fig. 2e) and proximity sensing (Fig. 2f). Additionally, IMU-based collision and vibration detection further enhance the system's multimodal sensing capabilities.

Performance characterization

To evaluate force and position sensing capabilities, we developed a testing platform incorporating an ATI Gamma sensor as the ground truth for force measurements (Fig. 3a). The evaluation utilized 48 probe (Supplementary Fig. 16) designs across three geometries (U, V and polygon shapes), collecting approximately 1,800 datasets per probe (Fig. 3b). A force sensing neural network (Fig. 3c) was developed based on a U-Net architecture[38], with ResNet48 (ref. 39) as the encoder to extract features from RGB deformation images captured by the sensor. A fully connected layer was added to output the resultant force vector, and the U-Net decoder generated a deformation mask. The mask was multiplied by the resultant vector to produce a force distribution map. The network was trained and evaluated using 86,440 sets of deformation data collected from 48 probe types (Fig. 3d), with a uniform sampling method employed to ensure comprehensive coverage of the sensor surface and accurately assess its force sensing performance. The dataset was split into 70% for training and 30% for testing. Training was conducted on a NVIDIA A6000 graphics processing unit using the L1 loss function and the AdamW optimizer, with a CosineAnnealingLR scheduler. The network achieves a position detection mean squared error accuracy of 0.056 mm and a 3D force detection mean squared error accuracy of 0.0004 N, with an overall position detection precision of around 0.4 mm (Fig. 3e) and a force error distribution of approximately 0.06 N (Fig. 3f), demonstrating robust performance across all probe types and strong generalizability (Supplementary Fig. 17). In addition, we conducted comparative experiments using ultraviolet and NIR modalities over 80,000 contact events to evaluate force sensing accuracy. The results showed that NIR consistently outperformed ultraviolet markers across all evaluation metrics, confirming its superior accuracy and stability in force sensing tasks (Supplementary Fig. 18). For 3D reconstruction testing, we not only optimized the distribution of markers in simulations but also evaluated the reconstruction accuracy of different algorithms. Through testing, our proposed method achieved an average root mean squared error of 0.0892 and a mean absolute error of 0.0375 (Supplementary Note 6). For surface characterization, a long short-term memory (LSTM) algorithm (Supplementary Note 7 and Supplementary Fig. 19) processed 150 sets of sliding and non-sliding data, achieving 97% accuracy in sliding detection. Colour classification was evaluated across six different colours, achieving 100% accuracy. Texture recognition was tested on six 3D-printed textures (Supplementary Fig. 20) and six common textures (Supplementary Fig. 21), demonstrating 98% accuracy (Fig. 3g,j). Additionally, the sensor exhibited robust capabilities in Braille sensing, as well as the perception of 0.07-mm-thick hair strands (Supplementary Fig. 22). To verify the accuracy of Braille recognition, we collected 200 samples for each of the 26 Braille letters, achieving a classification accuracy of 100%, which demonstrates the sensor's exceptional texture sensing capabilities (Supplementary Fig. 23).

Temperature detection was validated across a range of 0–90 °C, limited by the thermal resistance of the TPU film (Supplementary Figs. 24 and 25 and Supplementary Videos 4 and 5). The SuperTac can achieve a temperature sensing accuracy of 0.25 °C after testing and calibration and remains unaffected by ambient temperature variations within the 28–50 °C range. Heating induced by ultraviolet radiation causes only a minimal surface temperature change of 0.2 °C, ensuring negligible interference with MIR-based temperature measurements (Supplementary Note 8 and Supplementary Fig. 26).

The triboelectric sensing capability of SuperTac was comprehensively evaluated under diverse conditions, including ten different materials, seven contact-surface geometries, 15 contact speeds, three contact angles and five pressure levels (Supplementary Note 9 and Supplementary Figs. 27–29). Controlled experiments demonstrated robust classification performance in all situations, achieving 97% accuracy for contact angles, 99% accuracy for pressure levels, 96% accuracy for velocities and 95% accuracy for contact shapes, with an overall 95% accuracy across all conditions (Fig. 3k). A triboelectric signal acquisition platform was developed (Supplementary Fig. 30) to facilitate detailed signal analysis, and a 3.8 h durability test revealed consistently stable signal output (Supplementary Fig. 31). Furthermore, by employing advanced signal-filtering techniques and neural network classification, the triboelectric signals enabled proximity sensing within a range of 0–15 cm, depending on the material properties, underscoring the versatility and reliability of SuperTac in diverse sensing applications.

Vibration detection capabilities were validated using a custom platform (Supplementary Fig. 32), demonstrating accurate frequency recognition within the range of 0–60 Hz (Fig. 3i and Supplementary Fig. 33). For collision detection, we analysed 150 sets of IMU signals from collision and non-collision scenarios, achieving 94% classification accuracy (Fig. 3l and Supplementary Fig. 34).

Integration and applications

Robotic hand implementation

To demonstrate SuperTac's capabilities, we integrated it into two robotic platforms: a three-finger dexterous hand and a parallel gripper (Supplementary Video 6 and Supplementary Figs. 35 and 36). The dexterous hand features ten degrees of freedom with servo motor actuation at each joint. SuperTac is mounted in the palm, enabling comprehensive object-property sensing during grasping operations. For the parallel gripper configuration, SuperTac is installed on one side, facilitating stable object manipulation through integrated visual detection, contact force sensing, slip detection and collision detection algorithms.

Multimodal tactile language model

To enable advanced tactile information processing, we developed DOVE (Supplementary Note 10 and Supplementary Fig. 37), a multimodal tactile language model built on a pre-trained large language model (Fig. 4d). DOVE fuses multimodal tactile inputs and language to characterize object properties, reason over tactile differences between object pairs and infer an object's type and function. Specifically, DOVE can process triboelectric, temperature, colour and texture inputs to generate rich descriptions such as "yellow, room temperature, with a textured, raised, metallic surface" (Fig. 4d and Supplementary Video 7). When it receives tactile feedback from two objects, DOVE produces relational reasoning statements (for example, "The two objects share similar colours, temperatures and textures, but differ in material, so they are different."). DOVE also associates tactile impressions with semantic knowledge for reasoning (for example, "PET is commonly used for food containers. Its yellow colour suggests visibility or citrus-related items. This is probably a beverage bottle used for daily consumption."). To explore the impact of network structure on the perception capabilities of DOVE, we further investigated the effects of the hidden dimensions and activation functions in the projection layer. The experimental results demonstrated that changes in hidden dimensions had minimal impact on performance, whereas the Gaussian error linear

https://doi.org/10.1038/s44460-025-00006-y

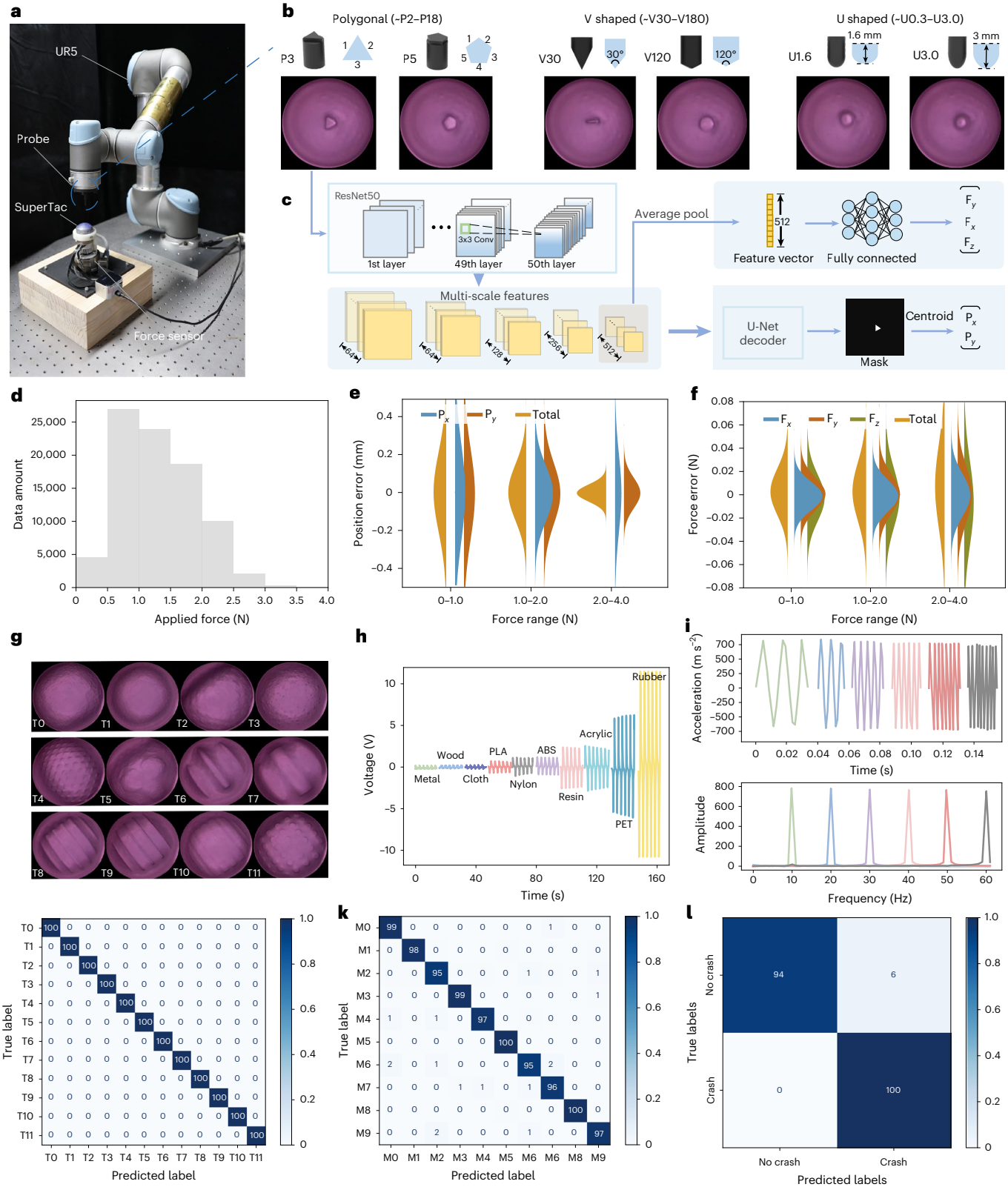

Fig. 3 | Perception and classification algorithm design. a, Force sensing data acquisition platform. **b**, We test the force sensing accuracy of 48 probes in U, V and polygonal shapes. **c**, Schematic of the force sensing network. 3x3 Conv, 3 × 3 convolutional layer.**d**, Graph showing the number of datasets in which various forces were applied. In total, we collected 86,440 datasets for contact force distribution. **e**, Contact position detection accuracy. **f**, Force sensing accuracy. **g**, Images showing the textures of 12 different surfaces. **h**, Triboelectric signals of ten different materials. **i**, Vibration signals at different frequencies are detected by the SuperTac. **j**, Texture classification confusion matrix. **k**, Material classification confusion matrix. **l**, Collision detection confusion matrix. ABS, acrylonitrile butadiene styrene; PET, polyethylene terephthalate; PLA, polylactic acid.

https://doi.org/10.1038/s44460-025-00006-y

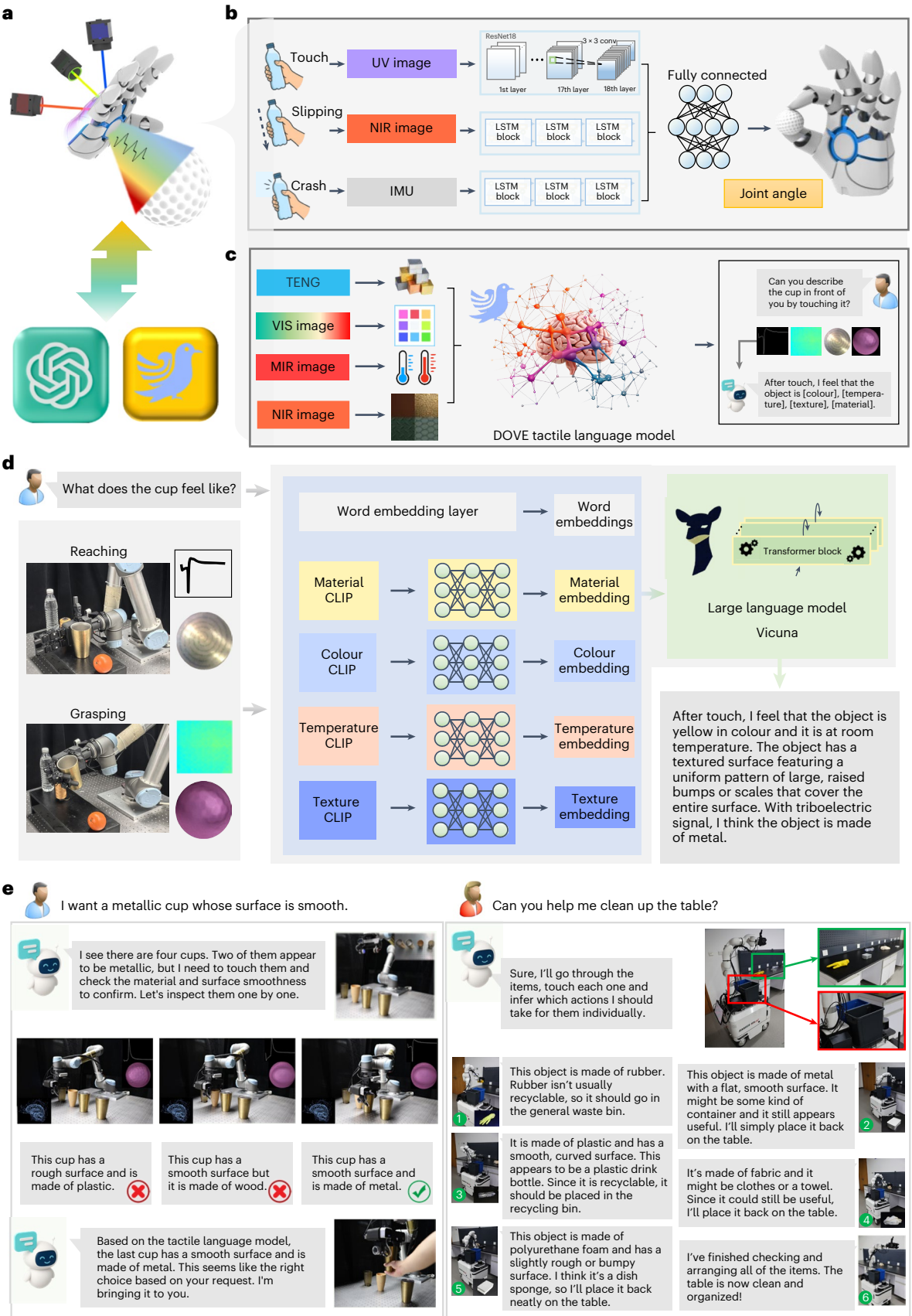

Fig. 4 | Design and application of the tactile language model. a, Schematic of the integration of SuperTac with DOVE in HRI. **b**, Stable object grasping by combining external vision with contact, slide and collision sensing. **c**, Fusion of material, texture, colour and temperature information, combined with a tactile language model for tactile information understanding. **d**, Schematic of our tactile language model and its application to tactile information understanding. **e**, Images and language from two experiments in HRI utilizing the tactile language model. The tactile language model assists robots in decision-making by providing detailed analyses and reasoning of tactile data. CLIP, contrastive language–image pre-training.

unit activation function notably outperformed the rectified linear unit function, ensuring effective alignment and fusion of multimodal features (Supplementary Note 11 and Supplementary Table 6).

Enhanced HRI

We further demonstrated the system's HRI capabilities across four experimental scenarios (Fig. 4e, Supplementary Notes 12–14, Supplementary Table 7, Supplementary Videos 8–12 and Supplementary Figs. 38 and 39). In the first scenario, the system identifies and selects a metallic cup with a smooth surface. In the second scenario, the system follows user instructions to locate a cup with specific characteristics (that is, lettering and a rough surface). GPT-4o orchestrates the interaction by directing visual identification and physical interaction with each cup, whereas DOVE processes the tactile feedback. In the third scenario, DOVE receives a reference object via touch and retrieves another that matches a specified colour by reasoning jointly over texture and colour cues. In the fourth scenario, DOVE infers cluttered tabletop objects' functions as reusable, recyclable or general waste based on tactile feedback and generates natural language justifications for each decision. The system continues evaluation until it finds a matching object or determines that no suitable matches exist.

The integration of comprehensive tactile sensing, language-based interpretation and visual processing represents a substantial advancement towards human-like robotic perception and interaction. By enabling robots to process and respond to multimodal sensory information in a manner akin to human perceptual capabilities, this approach paves the way for more intuitive and effective human–robot collaboration.

Conclusions

Traditional e-skin-based tactile sensors continue to face notable challenges in resolution, homogeneity and stability. Although visuotactile sensors offer promising solutions through advanced imaging techniques, their multimodal sensing capabilities have been limited by constraints in sensing skin design and imaging bandwidth. Our work addresses these fundamental limitations through a light-field-modulated sensing skin combined with multispectral imaging, enabling high-resolution multimodal sensing. The sensor achieves remarkable performance metrics, including 98% texture detection accuracy, 0.06 N 3D force detection accuracy in the NIR band, 97% sliding detection accuracy in the ultraviolet band and 100% colour detection accuracy in the visible band. By incorporating non-imaging perception inspired by pigeon magnetic field sensing, we further extend the sensor's capabilities to material detection (95% accuracy), collision detection (94% accuracy) and vibration detection (0–60 Hz range), all without compromising imaging quality or introducing electrode crosstalk issues.

The interpretation of heterogeneous tactile information through foundation models presents unique challenges. DOVE—our multimodal tactile language model—addresses these challenges through a unified input representation approach, which enhances scalability and adaptability across diverse sensor configurations. However, this approach reveals important trade-offs. Although transforming sequential data into images has proven effective for certain tasks, it may not fully capture the temporal characteristics inherent in tactile signals. Alternative approaches, such as time-series encoders, might better preserve temporal features but reintroduce challenges related to embedding heterogeneity. Striking the optimal balance between scalability and effectiveness remains a crucial area for future research and practical implementation.

Several promising directions have emerged for extending SuperTac's capabilities. Miniaturization of the sensor could enable fingertip installation, greatly advancing robotic in-hand manipulation capabilities. Additionally, DOVE's modality-agnostic framework, which converts various input modalities into image representations, could be adapted for different sensor configurations and applications. Future work will focus on advancing low-power decoding chips and exploring highly integrated packaging solutions to further reduce the sensor's size while addressing challenges in heat dissipation and system stability. Work will additionally focus on optimizing DOVE across diverse sensor designs and application-specific datasets, to enhance its versatility and robustness. These developments aim to bridge the gap between robotic and human-like perception capabilities, paving the way for more intuitive and effective HRI.

Methods

All of the experiments were conducted using Python 3.8.20 in a Conda environment. All of the analyses were performed on Ubuntu 20.04 with four NVIDIA RTX A6000 graphics processing units (CUDA 11.3).

Fabrication of the sensing skin

The sensing skin was fabricated using a multi-step process (Supplementary Fig. 40, Supplementary Note 15 and Supplementary Table 8). First, transparent silicone was mixed and poured into acrylic moulds, which provided a smoother surface finish compared with 3D-printed moulds. After heating, the silicone was cured to form the supporting layer. For the fluorescent layer, a scraping method was employed, using a steel mesh as a mask to spread fluorescent ink over the surface. To prevent unevenness caused by ink buildup, an additional layer of transparent silicone was applied using spin-coating. The reflective layer was created by mixing silver powder with transparent silicone, which was then spin-coated onto the fluorescent layer. For the conductive layer, conductive ink was screen printed onto a TPU surface and heated for 60 min to complete the layer. Finally, the conductive layer was attached to the translucent layer, finalizing the sensing skin. Although the integration of fluorescent markers introduces additional complexity, the standardized design ensures low cost (less than US$1) and high durability. The outer sensing skin, made of TPU film, which is commonly used in automotive and smartphone protective applications, exhibits exceptional wear and corrosion resistance. Fluorescent markers showed no photobleaching after one week of continuous ultraviolet exposure, ensuring stability (Supplementary Fig. 41). These features demonstrate a thoughtful balance between functional enhancements and cost-effectiveness.

Assembly and connection of SuperTac

The sensor was designed with a modular structure (Supplementary Note 16 and Supplementary Fig. 42) divided into three sections: upper, middle and lower. The upper and lower sections are made of aluminium alloy for high heat resistance and mechanical strength, whereas the middle section is constructed from transparent acrylic to ensure even diffusion of LED light onto the sensing skin. Threaded joints are used to connect the modules, allowing for easy disassembly. To address potential overheating during prolonged full-load operation, the SuperTac system incorporates a detachable magnetic cooling fan powered via contact-based pogo pins and aligned using N52-grade magnets, enabling quick removal for maintenance, and effectively reducing the stabilized temperature by 18.4 °C during extended high-load operation, as demonstrated through time–temperature comparison experiments (Supplementary Note 17 and Supplementary Figs. 43–46).

The SuperTac system adopts a USB 3.1 Gen 1 protocol for data communication, facilitating robust and high-speed transmission across all sensing and communication modules (Supplementary Note 18 and Supplementary Fig. 47). To ensure stable operation, the system is equipped with an optimized power architecture that supports all modules under full-load conditions, with a maximum power consumption of 4.5 W (Supplementary Note 19 and Supplementary Figs. 48 and 49). These design choices enhance the practicality and scalability of the SuperTac system in real-world applications. In addition, we designed a user interface that simultaneously displays signals including MIR, NIR, visible and ultraviolet light, triboelectric signals, posture information and acceleration data (Supplementary Fig. 50).

Image classification network design and training

For image-based tactile inputs, a ResNet18 backbone was cascaded with a multilayer perceptron (MLP) to extract task-relevant features and perform classification. The model processed batches of 128 × 128 visuotactile images, generating intermediate feature maps through ResNet, which were further processed via max-pooling and passed through the MLP classifier. The network was trained end to end for four tasks: colour, texture, temperature and material classification. Triboelectric signals were filtered to remove high-frequency components and visualized as curves, which were stored as images. The dataset was split into 80% for training, 10% for validation and 10% for testing. The model was trained using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 128, alongside a step scheduler that reduced the learning rate by 0.9 every ten validation steps.

Sequential signal classification network design and training

For sequential inputs (for example, IMU data and visuotactile videos), an LSTM network was employed as the backbone to process the temporal flow of information (Supplementary Fig. 19). Low-dimensional data, such as IMU readings, were processed using a two-layer MLP, whereas spatial–structural data, such as videos, were processed using a pre-trained ResNet18. The LSTM updated its hidden state sequentially and output task-oriented information, which was passed through an MLP classifier for final prediction. For IMU data, the model was trained end to end for collision detection, whereas for sliding detection, only the LSTM and MLP classifiers were trained. The dataset was split into 80% for training, 10% for validation and 10% for testing. Training used the AdamW optimizer with a learning rate of 1×10^{-3} and a batch size of 128, alongside a step-based learning-rate scheduler. After testing, the classification algorithms based on ResNet and LSTM have a single prediction time within 6 ms, meeting real-time requirements (Supplementary Note 20 and Supplementary Table 9).

Effects of air pressure and object hardness on sensor perception

We investigated the impact of internal air pressure on the tactile sensing performance of the sensor, focusing on its ability to perceive flexible objects and its accuracy in force sensing, texture recognition and sliding detection. During testing, five pressure levels (1.2, 3.0, 4.0, 6.0 and 7.0 kPa) were selected for the force sensing experiments, whereas three pressure levels (3.0, 5.0 and 7.0 kPa) were chosen for the texture recognition and sliding detection experiments. The experimental results demonstrated that variations in air pressure had minimal impact on the accuracy of force sensing, texture recognition and sliding detection. Notably, texture recognition and sliding detection achieved 100% accuracy across all pressure conditions. A slight decrease in force sensing accuracy was observed at high pressure (7 kPa), but it remained within an acceptable range. Overall, the system exhibited stable and reliable performance under varying pressure conditions (Supplementary Note 21 and Supplementary Figs. 51 and 52).

Extensive testing of the SuperTac system was conducted on soft and liquid-containing objects, including probes made of diverse materials (polylactic acid, cloth, plastic, paper, polyethylene terephthalate or silicone) and objects with flexible or liquid-containing textures. Although the softness of objects slightly impacted force sensing accuracy, the performance significantly improved after supplementing the dataset with 500 flexible object samples (Supplementary Note 22 and Supplementary Fig. 53). The system achieved 100% accuracy in texture recognition and sliding detection (Supplementary Figs. 54 and 55). Furthermore, the inflatable structure of SuperTac demonstrated superior texture and contour sensing capabilities compared with GelSight Mini, highlighting its advantages in handling complex surfaces (Supplementary Fig. 56). Additionally, simulation results using finite element analysis revealed that the system maintains reliable contour recognition for objects with elastic moduli above 1 MPa, providing theoretical guidance for practical applications (Supplementary Note 23 and Supplementary Fig. 57).

Tactile language model design and training

To enable comprehensive understanding and reasoning over multimodal tactile data and language, a large tactile language model was trained on a processed dataset integrating colour, texture, temperature and triboelectric data, augmented with synthetic tactile language question and answer pairs (Supplementary Figs. 37 and 39). The training and testing data for the SuperTac system were constructed using tactile data spanning six colours, three temperature conditions, ten material types and six surface textures, with multimodal question and answer pairs generated by GPT-4 and rule-based scripts to integrate tactile information with natural language descriptions (Supplementary Note 24). The training involved three stages: encoder pre-training, embedding alignment and model fine-tuning. Pre-trained contrastive language–image pre-training models[40] were used to extract image features, with an MLP classifier attached for end-to-end classification. After fine-tuning, the classifiers were removed and a projection layer was added for embedding alignment. Finally, the projection layer and language backbone (Vicuna[41]) were fine-tuned using LoRA[42]. The total parameters of the four contrastive language–image pre-training encoders and language backbone reached 8.6 billion. Training used the AdamW optimizer with a cosine annealing schedule, achieving robust performance across all modalities (Supplementary Note 25 and Supplementary Tables 10–12).

Software availability

All of the experiments were conducted using Python 3.8.20 in a Conda environment. All of the analyses were performed on Ubuntu 20.04 with four NVIDIA RTX A6000 graphics processing units (CUDA 11.3).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available at https://cloud.tsinghua.edu.cn/d/f6abfcf5845a42018e2a/files/?p=%2FData%2Fdataset.zip.

Code availability

We have open sourced the codebase for DOVE at https://github.com/wut19/DOVE. Future updates and new releases will also be available at this link.

References

1. Bauza, M. et al. SimPLE, a visuotactile method learned in simulation to precisely pick, localize, regrasp, and place objects. *Sci. Robot.* **9**, eadi8808 (2024).
2. Luo, Y. et al. Adaptive tactile interaction transfer via digitally embroidered smart gloves. *Nat. Commun.* **15**, 868 (2024).
3. Sun, Z., Zhu, M., Shan, X. & Lee, C. Augmented tactile–perception and haptic–feedback rings as human–machine interfaces aiming for immersive interactions. *Nat. Commun.* **13**, 5224 (2022).
4. Abad, A. C., Reid, D. & Ranasinghe, A. HaptiTemp: a next-generation thermosensitive GelSight-like visuotactile sensor. *IEEE Sens. J.* **22**, 2722–2734 (2022).
5. Zhang, N. et al. Soft robotic hand with tactile palm–finger coordination. *Nat. Commun.* **16**, 2395 (2025).
6. Zhong, D. et al. High-speed and large-scale intrinsically stretchable integrated circuits. *Nature* **627**, 313–320 (2024).
7. Qiu, Y. et al. Quantitative softness and texture bimodal haptic sensors for robotic clinical feature identification and intelligent picking. *Sci. Adv.* **10**, eadp0348 (2024).

8. Çeliker, H., Dehaene, W. & Myny, K. Multi-project wafers for flexible thin-film electronics by independent foundries. *Nature* **629**, 335–340 (2024).
9. Yuan, W., Dong, S. & Adelson, E. H. Gelsight: high-resolution robot tactile sensors for estimating geometry and force. *Sensors* **17**, 2762 (2017).
10. Yang, F. et al. Binding touch to everything: learning unified multimodal tactile representations. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 26340–26353 (IEEE, 2024).
11. Fu, L. et al. A touch, vision, and language dataset for multimodal alignment. In *Proc. International Conference on Machine Learning* 14080–14101 (PMLR, 2024).
12. Mu, S. et al. A platypus-inspired electro-mechanosensory finger for remote control and tactile sensing. *Nano Energy* **116**, 108790 (2023).
13. Hochstoeger, T. et al. The biophysical, molecular, and anatomical landscape of pigeon CRY4: a candidate light-based quantal magnetosensor. *Sci. Adv.* **6**, eabb9110 (2020).
14. Xu, J. et al. Magnetic sensitivity of cryptochrome 4 from a migratory songbird. *Nature* **594**, 535–540 (2021).
15. Shi, H., Liu, C., Jiang, Q. & Xu, J. Effective approaches to improve the electrical conductivity of PEDOT: PSS: a review. *Adv. Electron. Mater.* **1**, 1500017 (2015).
16. Sun, H., Kuchenbecker, K. J. & Martius, G. A soft thumb-sized vision-based sensor with accurate all-round force perception. *Nat. Mach. Intell.* **4**, 135–145 (2022).
17. Hogan, F. R. et al. Seeing through your skin: recognizing objects with a novel visuotactile sensor. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision* 1218–1227 (IEEE, 2021).
18. Li, S. et al. M^3Tac: a multispectral multimodal visuotactile sensor with beyond-human sensory capabilities. *IEEE Trans. Rob.* **40**, 4506–4525 (2024).
19. Park, K. et al. A biomimetic elastomeric robot skin using electrical impedance and acoustic tomography for tactile sensing. *Sci. Robot.* **7**, eabm7187 (2022).
20. Zhang, Y., Kan, Z., Tse, Y. A., Yang, Y. & Wang, M. Y. Fingervision tactile sensor design and slip detection using convolutional lstm network. Preprint at https://arxiv.org/abs/1810.02653 (2018).
21. Ward-Cherrier, B. et al. The TacTip family: soft optical tactile sensors with 3D-printed biomimetic morphologies. *Soft Robot.* **5**, 216–227 (2018).
22. Ge, J. et al. A bimodal soft electronic skin for tactile and touchless interaction in real time. *Nat. Commun.* **10**, 4405 (2019).
23. Massari, L. et al. Functional mimicry of Ruffini receptors with fibre Bragg gratings and deep neural networks enables a bio-inspired large-area tactile-sensitive skin. *Nat. Mach. Intell.* **4**, 425–435 (2022).
24. Liu, W. et al. Touchless interactive teaching of soft robots through flexible bimodal sensory interfaces. *Nat. Commun.* **13**, 5030 (2022).
25. Jiang, Y. et al. A multifunctional tactile sensory system for robotic intelligent identification and manipulation perception. *Adv. Sci.* **11**, 2402705 (2024).
26. Mao, Q., Liao, Z., Yuan, J. & Zhu, R. Multimodal tactile sensing fused with vision for dexterous robotic housekeeping. *Nat. Commun.* **15**, 6871 (2024).
27. Wang, Y. et al. Hierarchically patterned self-powered sensors for multifunctional tactile sensing. *Sci. Adv.* **6**, eabb9083 (2020).
28. Cai, M. et al. A multifunctional electronic skin based on patterned metal films for tactile sensing with a broad linear response range. *Sci. Adv.* **7**, eabl8313 (2021).
29. Hua, Q. et al. Skin-inspired highly stretchable and conformable matrix networks for multifunctional sensing. *Nat. Commun.* **9**, 244 (2018).
30. Kim, K. et al. Extremely durable electrical impedance tomography-based soft and ultrathin wearable e-skin for three-dimensional tactile interfaces. *Sci. Adv.* **10**, eadr1099 (2024).
31. Tian, X. et al. High-resolution carbon-based tactile sensor array for dynamic pulse imaging. *Adv. Funct. Mater.* **34**, 2406022 (2024).
32. Roorda, A. & Williams, D. R. The arrangement of the three cone classes in the living human eye. *Nature* **397**, 520–522 (1999).
33. Mouritsen, H. & Hore, P. J. The magnetic retina: light-dependent and trigeminal magnetoreception in migratory birds. *Curr. Opin. Neurobiol.* **22**, 343–352 (2012).
34. Li, S. et al. When vision meets touch: a contemporary review for visuotactile sensors from the signal processing perspective. *IEEE J. Sel. Top. Signal Process.* **18**, 267–287 (2024).
35. Li, S. et al. Visual-tactile fusion for transparent object grasping in complex backgrounds. *IEEE Trans. Rob.* **39**, 3838–3856 (2023).
36. Suresh, S. et al. NeuralFeels with neural fields: visuotactile perception for in-hand manipulation. *Sci. Robot.* **9**, eadl0628 (2024).
37. Brohan, A. et al. Rt-2: vision–language–action models transfer web knowledge to robotic control. In *Proc. Conference on Robot Learning* 2165–2183 (PMLR, 2023).
38. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *Proc. Medical Image Computing and Computer-Assisted Intervention* 234–241 (Springer, 2015).
39. 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).
40. Radford, A. et al. Learning transferable visual models from natural language supervision. In *Proc. International Conference on Machine Learning* 8748–8763 (PMLR, 2021).
41. Zheng, L. et al. Judging LLM-as-a-judge with MT-bench and Chatbot Arena. In *Proc. Advances in Neural Information Processing Systems* 46595–46623 (Curran Associates, 2023).
42. Hu, E. J. et al. LoRA: low-rank adaptation of large language models. In *Proc. of International Conference on Learning Representations* 12513–12525 (OpenReview.net, 2022).

Acknowledgements

Funding was provided by the National Key Research and Development Program of China (2024YFB3816000 to W.D.), Shenzhen Science and Technology Program (KJZD20240903100905008 to W.D.), Guangdong Innovative and Entrepreneurial Research Team Program (2021ZT09L197 to W.D.), Tsinghua Shenzhen International Graduate School–Shenzhen Pengrui Young Faculty Program of Shenzhen Pengrui Foundation (SZPR2023005 to W.D.), National Natural Science Foundation of China (12472160 to Z.X.), Research Grants Council of the Hong Kong Special Administrative Region (RFS2324-1S03, R1017-24F, T42-513/24-R, C7005-23Y, 11211425, 11215722 and 11211523 to X.Y.), National University of Singapore Presidential Young Professorship Start-Up Grant (A-8000380-02-00 to C.W.), Singapore Ministry of Education Academic Research Fund (A-8002146-00-00 to C.W.) and Tsinghua–NUS Joint Research Initiative Fund Award (A-8002542-00-00 to C.W.).

Author contributions

W.D. and S.L. conceived of the idea and guided the project. S.L., T.W. and J.X. designed the experiments, analysed the data and drafted the manuscript. Z.X., C.W. and X.Y. instructed on manuscript writing and experimentation. S.L. and T.W. performed the characterization of

the material. S.L. and T.W. conducted functional experiments on the sensor. T.W. and Y.H. designed and implemented the classification algorithms. S.L. and J.X. contributed to the mechanical design. Z.Z., H.Z. and Y.Y. conducted the theoretical analysis and simulations. S.L., T.W., Q.X., Z.W., S.M., L.Y., X.W., Z.X., C.L., C.W., X.Y. and W.D. revised the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44460-025-00006-y.

Correspondence and requests for materials should be addressed to Zhaoqian Xie, Changsheng Wu, Xinge Yu or Wenbo Ding.

Peer review information *Nature Sensors* thanks Chengkuo Lee and the other, anonymous, reviewers for their contribution to the peer review of this work.

[1]Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. [2]State Key Laboratory of Structural Analysis, Optimization and CAE Software for Industrial Equipment, Dalian University of Technology, Dalian, China. [3]Department of Engineering Mechanics, Dalian University of Technology, Dalian, China. [4]Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA, USA. [5]School of Future Technology, Shanghai University, Shanghai, China. [6]Department of Mechanics and Engineering Science, Sichuan University, Chengdu, China. [7]Xspark AI, Shenzhen, China. [8]Department of Materials Science and Engineering, National University of Singapore, Singapore, Singapore. [9]Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore. [10]Institute for Health Innovation and Technology, National University of Singapore, Singapore, Singapore. [11]The N.1 Institute for Health, National University of Singapore, Singapore, Singapore. [12]Department of Biomedical Engineering, City University of Hong Kong, Hong Kong, China. [13]Institute of Digital Medicine, City University of Hong Kong, Hong Kong, China. [14]Hong Kong Institute for Clean Energy, City University of Hong Kong, Hong Kong, China. [15]These authors contributed equally: Shoujie Li, Tong Wu, Jianle Xu. ✉e-mail: zxie@dlut.edu.cn; cwu@nus.edu.sg; xingeyu@cityu.edu.hk; ding.wenbo@sz.tsinghua.edu.cn

nature portfolio

Corresponding author(s): Xinge Yu

Last updated by author(s): Oct 16, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Custom Python 3.8 code was used to collect the data in this study.
Data analysis	Custom Python 3.8 code was utilized for data analysis in this study, while the finite element analysis (FEA) software ABAQUS 2019 was employed to simulate the mechanical response of the tactile sensor. Additional packages used in python environment are list as follows: accelerate==1.0.1, bitsandbytes==0.45.5, certifi==2025.4.26, charset-normalizer==3.4.2, filelock==3.16.1, fsspec==2025.3.0, hf-xet==1.1.2, huggingface-hub==0.32.3, idna==3.10, Jinja2==3.1.6, MarkupSafe==2.1.5, mpmath==1.3.0, natsort==8.4.0, networkx==3.1, numpy==1.24.4, nvidia-cublas-cu12==12.1.3.1, nvidia-cuda-cupti-cu12==12.1.105, nvidia-cuda-nvrtc-cu12==12.1.105, nvidia-cuda-runtime-cu12==12.1.105, nvidia-cudnn-cu12==9.1.0.70, nvidia-cufft-cu12==11.0.2.54, nvidia-curand-cu12==10.3.2.106, nvidia-cusolver-cu12==11.4.5.107, nvidia-cusparse-cu12==12.1.0.106, nvidia-nccl-cu12==2.20.5, nvidia-nvjitlink-cu12==12.9.41, nvidia-nvtx-cu12==12.1.105, opencv-python==4.11.0.86, packaging==25.0, peft==0.2.0, pillow==10.4.0, protobuf==5.29.5, psutil==7.0.0, python-dotenv==1.0.1, PyYAML==6.0.2, regex==2024.11.6, requests==2.32.3, safetensors==0.5.3, sentencepiece==0.2.0, sympy==1.13.3, tokenizers==0.13.3, torch==2.4.1, torchvision==0.19.1, tqdm==4.67.1, transformers==4.31.0, triton==3.0.0, typing_extensions==4.13.2, urllib3==2.2.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Codes used for training and evaluating DOVE, the proposed tactile language model, are publicly available at https://github.com/wut19/DOVE .The corresponding multimodal tactile-language dataset supporting DOVE is available at https://cloud.tsinghua.edu.cn/d/f6abfcf5845a42018e2a/files/?p=%2FData%2Fdataset.zip. Codes and datasets used in other experimental categories (e.g., classification or comparative baselines) are not publicly released due to institutional and third-party restrictions, but are available from the corresponding author upon reasonable request.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	N/A. No human participants, human data, or human biological materials were involved in this study.
Reporting on race, ethnicity, or other socially relevant groupings	N/A. No human participants, human data, or human biological materials were involved in this study.
Population characteristics	N/A. No human participants, human data, or human biological materials were involved in this study.
Recruitment	N/A. No human participants, human data, or human biological materials were involved in this study.
Ethics oversight	N/A. No human participants, human data, or human biological materials were involved in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	To ensure that the data volume met the experimental requirements, we collected 2,160, 2,250, 2,160, and 10,500 raw samples for color, temperature, texture and material, respectively, yielding a combinatorial space of 1×10^14 unique instances. We trained DOVE on 30,000 synthetic combinations and evaluated it on newly generated 1000 samples.
Data exclusions	No data were excluded from analysis.
Replication	All attempts at replication were successful.
Randomization	No human or animal participants were involved; the study concerns a tactile sensor evaluated with different probes and test objects. Randomization was therefore not relevant to the study.
Blinding	No human or animal participants were involved; therefore, conventional investigator blinding to participant group allocation is not applicable. Data acquisition was automated via predefined scripts that controlled probe motion, normal force targets, and sampling timing.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A