RESEARCH ARTICLE

ADVANCED INTELLIGENT SYSTEMS Open Access

www.advintellsyst.com

TacEva: A Performance Evaluation Framework for Vision-Based Tactile Sensors

*Qingzheng Cong, Steven Oh, Wen Fan, Shan Luo, Kaspar Althoefer, and Dandan Zhang**

Vision-based tactile sensors (VBTSs) are widely used in robotic tasks because of the high spatial resolution and relatively low manufacturing costs they offer. However, variations in their sensing mechanisms, structural dimensions, and other parameters lead to significant performance disparities between existing VBTSs. This makes it challenging to optimize them for specific tasks, as both the initial choice and subsequent fine-tuning are hindered by the lack of standardized metrics. To address this issue, TacEva is introduced as a comprehensive evaluation framework for the quantitative analysis of VBTS performance. This framework defines a set of performance metrics that capture key characteristics in typical application scenarios. For each metric, a structured experimental pipeline is designed to ensure consistent and repeatable quantification. TacEva has been applied to multiple VBTSs with distinct sensing mechanisms, and the results demonstrate its ability to provide a thorough evaluation of each design and quantitative indicators for each performance dimension. This enables researchers to preselect the most appropriate VBTS on a task-by-application basis, while also offering performance-guided insights into the optimization of VBTS design. A list of existing VBTS evaluation methods and additional evaluations can be found on our website: https://stevenoh2003.github.io/TacEva/.

1. Introduction

Robots have yet to attain the level of manipulative dexterity exhibited by humans, a challenge rooted in the difficulty of accurately acquiring detailed contact information in physical environments.[1] Tactile sensing has therefore become indispensable for delicate and precise robotic manipulation in embodied intelligence systems.[2] A notable development in this domain has been the rise of vision-based tactile sensors (VBTSs).[3] These sensors employ high-resolution cameras to capture detailed contact surface information, thereby integrating seamlessly with computer vision and image-based deep learning methods. Nevertheless, among VBTSs, we see a wide variety of architecture designs, structural dimensions, sensing mechanisms, and fabrication techniques, depending on the exact nature of the application requirements.[4]

Such a rapid development of VBTSs has created an urgent need for standardized performance evaluation. Selecting an appropriate VBTS for a specific task scenario remains challenging, as distinct sensor designs offer substantial variation in performance. A universal evaluation protocol would therefore support informed selection, guide design optimization, and facilitate fair comparison across sensor designs. However, a significant gap persists in the field: (1) there is currently no standardized framework for the evaluation of VBTS, (2) the inconsistency of current metrics limits the objective, reproducible, and comprehensive assessment of the multiple sensors, (3) the above two challenges are further compounded by the inherently multimodal nature of tactile sensing[5] from spatial fine texture and geometry contact depth, dynamic contact torque and distributed force, toward temporal vibration frequency and slip response, which requires coordinated evaluation across multiple performance dimensions.

In detail, there are numerous reasons that lead to the aforementioned research gap: (1) The foremost reason is that the

---

Q. Cong, W. Fan, D. Zhang
Department of Bioengineering
Imperial-X
Imperial College London
London W12 0BZ, UK
E-mail: d.zhang17@imperial.ac.uk

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202501179.

DOI: 10.1002/aisy.202501179

S. Oh
Department of Modern Mechanical Engineering
Waseda University
Tokyo 169-8555, Japan

S. Luo
Department of Engineering
King's College London
London, UK

K. Althoefer
Centre for Advanced Robotics
School of Engineering and Materials Science
Queen Mary University of London
London, UK

application scenarios for VBTSs are inherently diverse. This makes it difficult to establish universal and broadly applicable evaluation metrics that remain meaningful across the field. (2) Further complexity arises from VBTS hardware that is commonly fabricated using elastomers, light-emitting diode (LED) illuminations, and cameras, thereby introducing additional considerations related to soft-material mechanics, optics, and imaging. (3) Moreover, owing to different sensing mechanisms, each VBTS design inevitably exhibits its own profile of strengths and weaknesses. However, certain positive performance characteristics will tend to be emphasized, while weaker ones may be underreported.

To address such a research gap, we propose TacEva, a systematic evaluation framework that integrates quantitative performance metrics with a structured and reproducible assessment pipeline applicable to a wide range of VBTS designs. From **Figure 1**, TacEva organizes evaluation into three categories: (1) Intrinsic metrics: capturing the physical and optical fundamental parameters of VBTS designs that influence their tactile sensing representations. (2) Standard performance: quantifying the inference fidelity from tactile sensing representations into tactile perception values under reproducible contact stimuli, reflecting both tactile sensing quality and the effectiveness of the perception model. (3) Robustness: characterizing the stability of tactile perception outputs when the input domain of tactile sensing changes due to variations in contact location, illumination, or repeated trials.

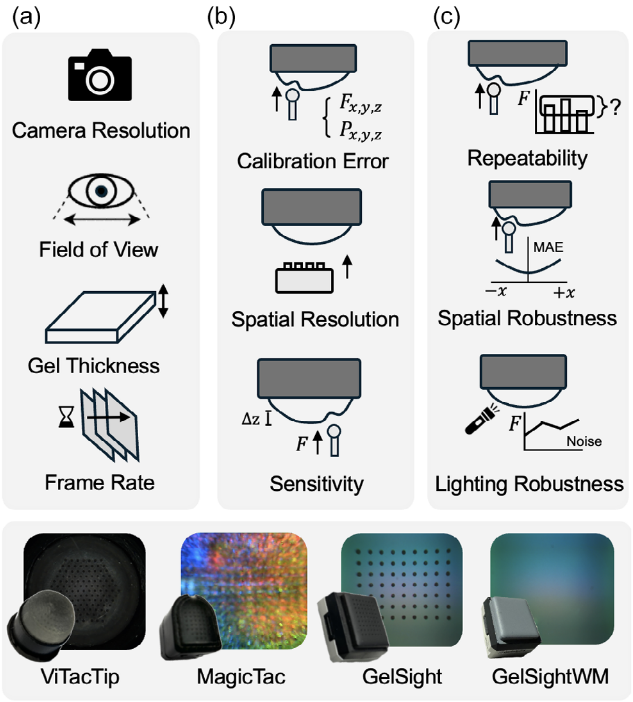

Figure 1. TacEva benchmarking pipeline. A unified framework for evaluating VBTSs that enables systematic comparison across a) intrinsic metrics, b) standard performance, and c) robustness. We demonstrate TacEva on four representative VBTSs: ViTacTip,[11] MagicTac,[41] GelSight Mini Tracking Marker Gel (GelSight), and GelSight Mini Standard Silicone Gel (GelSightWM).[3]

Taken together, these three dimensions give TacEva a broadly applicable and sensor-agnostic structure without redesigning task-specific metrics. They mirror the main factors that determine real-world VBTS performance: (1) Hardware and optical constraints on tactile sensing (intrinsic metrics), (2) inference quality under predefined conditions (standard performance), and (3) tactile perception stability under domain shifts (robustness). Whereas existing VBTS evaluations typically probe only one or two of these aspects in isolation.

In summary, TacEva provides a unified framework for consistent and comparable assessment of VBTSs. Defining and standardizing a comprehensive set of performance metrics and evaluation protocols enables practitioners to make evidence-based sensor choices and offers sensor developers clear guidance for targeted optimization and comparison across heterogeneous designs.

2. Related Work

2.1. Development of VBTSs

Several works have reviewed and summarized the development of VBTSs from the perspectives of structural design[4] and sensing mechanisms[6] and have also demonstrated their usability in diverse robotic applications.[7] Building on these reviews, the CrystalTac[8] offers a systematic classification of VBTSs according to their sensing principles. These include: (i) Intensity mapping method (IMM), which infers contact geometry and pressure through spatial variations in pixel value intensity; (ii) marker displacement method (MDM), which detects surface deformation by tracking the displacement of embedded markers under force; and (iii) modality fusion method (MFM), which introduces extra sensing modalities beyond tactile only to enable multimodal perception.

GelSight-type sensors[3] are IMM-based and are widely used for reconstructing fine contact geometry, enabled by a reflective coating on a compliant surface. Later updates also incorporate markers on the gel surface, thereby combining both IMM with MDM to improve deformation tracking and force regression. Based on IMM–MDM design, the F-TOUCH sensor[9] introduces a spring-mechanism structure for accurate force/torque estimation within a compact device. TacTip-type sensors are MDM-based and capture dynamic features such as force, torque, and vibration via marker geometric patterns.[10] Recent work increasingly explores MFM-based sensors, such as those that combine vision, proximity, or thermal with tactile sensing,[11–13] to enhance versatility. However, as VBTS designs become more complex, so does their evaluation, underscoring the need for standardized assessment frameworks.

2.2. Evaluation of Tactile Sensors

Existing metrics to evaluate tactile sensors typically include design-level, performance-level, and task-level. For design-level metrics, sensing area, structural geometry, and compactness play an essential role as intrinsic features of sensor designs. At the performance level, metrics examine spatial sensing density, depth reconstruction accuracy, force/torque dynamic sensing range, contact durability, and so on. Finally, at the task level, downstream performance is measured through application-driven metrics such as reconstruction error or manipulation success rate, which directly reflect how well the overall tactile system supports a given robotic task.[14,15]

2.2.1. Non-VBTSs

Non-VBTSs, such as magnetic (uSkin,[16] Hall-effect arrays[17]), capacitive (iCub fingertips,[18] PPS TactArray[19]), piezoresistive (Weiss WTS arrays[20]), piezoelectric (PVDF-based skins[21]), impedance/fluidic (BioTac[22]), and barometric/fluidic (BaroTac[23]), all have established, well-defined evaluation practices. The following examples illustrate this. uSkin is evaluated in terms of its precalibration response, including axis crosstalk, overload durability, calibration accuracy (via R^2 comparisons across fitting models), hysteresis during load–unload cycles, and signal-to-noise ratio (SNR).[16] Capacitive fingertips are typically assessed by pre-/postmolding SNR, accumulated-load nonlinearity and hysteresis, temperature-drift compensation, and full triaxial calibration.[24] Piezoelectric skins report measurable force range, axial sensitivity, nonlinearity, repeatability, frequency response, and mechanical flexibility.[21,25] Recent custom designs also quantify force-prediction accuracy and robustness to magnetic noise.[26] As can be seen, non-VBTSs benefit from relatively standardized design- and performance-level evaluation protocols, consistent with broader efforts in robotics to establish tactile performance benchmarking frameworks.[27]

2.2.2. VBTSs

Evaluating VBTS remains challenging because evaluation methods vary widely across studies. Individual prototypes such as GelForce,[28] FingerVision,[12] DIGIT,[29] OmniTact,[30] DenseTact,[31] ViTacTip,[11] ThinTact,[32] and many others are typically validated on self-chosen tasks, with different stimuli, tooling, and learning pipelines. As a result, their reported metrics cover different mixtures of tasks, including contact-shape reconstruction, contour tracking, force estimation, grasping-force control, and robustness tests such as long-term durability, while intrinsic properties and robustness are often only partially documented. while robustness is often only partially documented.

Broader cross-sensor evaluation efforts are in two main directions. First, hardware-focused works report design-level intrinsic metrics (e.g., sensing area, optical layout, gel configuration) and discuss how they influence geometry and force sensing.[13,33–36] In parallel, task-focused efforts such as Eval[37] and Sparsh[38] propose application-level benchmarks and compare downstream performance on representative tasks (e.g., edge detection, dexterous manipulation, multitask representation learning). Both directions substantially advance VBTS development, yet each of which alone has constraints, as they address complementary aspects: (1) Intrinsic metrics alone do not reveal how a sensor design behaves in downstream tasks, (2) while task-driven metrics are tied closely to sensor designs, experimental setup, or algorithmic choices, leading to less generalisability across sensors. Moreover, existing evaluation protocols differ in their metric definitions, calibration procedures, and experimental setups, and thus do not offer a sensor-agnostic, performance-level middle layer of metrics evaluation that can be easily reproduced.

To quantify this inconsistency, we surveyed representative VBTSs and evaluation-oriented works. Table A1 summarizes which intrinsic, standard-performance, and robustness metrics are reported in each article. In particular, we noticed that repeatability and robustness are rarely documented, highlighting the need for a unified and reproducible evaluation framework.

To bridge this gap, we introduce TacEva, a unified framework that: (i) provides a comprehensive evaluation framework for VBTSs by covering the three fundamental aspects of VBTS performance, intrinsic hardware metrics, standard operating performance, and robustness; (ii) formalizes sensor-agnostic experimental protocols so that the same metric definitions and procedures can be applied, unchanged, to heterogeneous VBTS architectures; (iii) provides a shared performance-level basis that links low-level sensor characteristics to typical task requirements, offering a practical decision aid for selecting and comparing VBTSs in downstream robotic applications.

3. TacEva Framework Overview

TacEva is designed to be generalizable across VBTSs' architectures. The framework only requires a VBTS mounted on a robotic arm and an external six-axis force–torque sensor for ground-truth measurement, and does not rely on specific sensor geometries, marker layouts, elastomer materials, or perception models. All remaining components of the setup can be reproduced through simple 3D printing. The same protocol and metric definitions can therefore be used on any existing or future VBTS design.

The pipeline evaluates VBTS from three primary aspects: (1) intrinsic metrics (Figure 1a), (2) standard performance (Figure 1b), and (3) robustness (Figure 1c). Section 3.1 formalizes the intrinsic properties. Section 3.2–3.4 detail the experimental protocol and metrics for standard performance. Last, Section 3.5 and 3.6. Section 3.7 defines the robustness criteria and describes the corresponding experimental procedures.

3.1. Intrinsic Properties

3.1.1. Definitions

Properties intrinsic to the hardware parameters of a specific VBTS design, such as the camera optics and elastomer geometry, can be regarded as intrinsic metrics when compared to each other: 1) Camera resolution: The pixel resolution of the internal camera CMOS chip. Under the same design of sensing mechanism, the higher camera resolution captures finer raw tactile image with richer contact information, especially for IMM designs[3] which directly rely upon pixel value. 2) Gel thickness: The thickness (in mm) of the compliant elastomer layer, typically measured at the dome's center. A thicker gel increases compliance and durability but may reduce spatial resolution (SR) and introduce optical distortion. 3) Field of view (FOV): The effective sensing area (in mm^2) of the tactile surface that is visible to the internal camera. A larger FOV allows coverage of a greater portion of the elastomer surface, improving the detection of distributed or large-area contact patterns. In this work, we report projected area rather than angular FOV, as it directly reflects the usable sensing region and enables comparison across different sensor geometries. 4) Frame rate (Frames Per Second): The image-capture frequency (in Hz) of the internal camera. Higher FPS improves temporal resolution to track fast or dynamic contact events.

These properties describe the intrinsic hardware characteristics of a VBTS, which set the fundamental limits on sensing quality and form the basis for evaluating higher-level performance.

3.2. Calibration

3.2.1. Definitions

Accurate characterization of a VBTS requires calibrating both its surface geometry and its mapping from raw optical outputs to contact forces and positions. This is achieved through two sequential procedures with the sensor mounted on a robotic manipulator: (1) *surface-geometry reconstruction*, which recovers the 3D shape of the sensor surface to localize contact points precisely; and (2) *force-localization calibration*, which maps the raw optical outputs of the sensor to ground-truth forces and indentation positions.

3.2.2. Experimental Protocols

To obtain these calibration results, a six-axis force/torque transducer (M3813B) is rigidly affixed to the workbench, and a 2-mm-radius, 3D-printed spherical indenter is attached to its sensing head. The VBTS is then mounted on a robot arm for data collection.

As shown in **Figure 2** (Stage 1), the robot arm moves the indenter toward the sensor surface, along the vertical axis (z-axis), in small steps of 0.1 mm until contact is detected. The corresponding end-effector position at the moment of initial contact is recorded as a point on the sensor surface. In Stage 2 of Figure 2, after each initial contact point detected in Stage 1, the robot performs four additional indentations to randomized depths z_n in the vertical direction. At each indentation depth, a small displacement is applied in the $x-y$ plane to introduce variation in the contact direction and to capture both normal and shear-force responses. Note that the maximum indentation depth and normal force were constrained by the safe operating

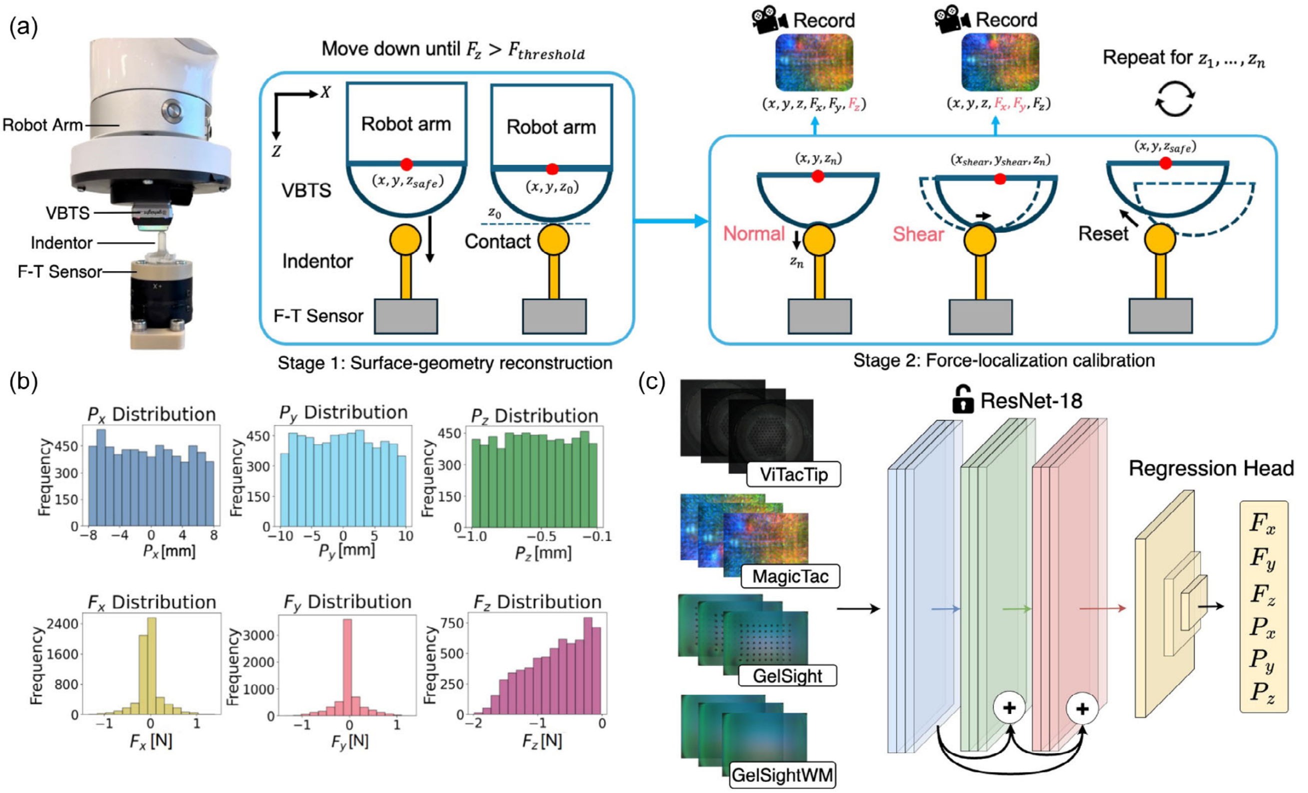

Figure 2. Calibration procedure. a) Experimental data collection. Tool positions (P_x, P_y) were uniformly sampled across the sensor surface at a fixed insertion depth P_z, then randomly perturbed along the x, y, and z axes to generate the dataset. Our framework implements a two-stage calibration procedure Stage 1. Surface-geometry reconstruction to map the 3D sensor surface using a robotic arm and force-sensitive spherical indenter (10mm radius), and Stage 2. Force-localization calibration establishes optical-output-to-force mappings through systematic probing. The two stages are done sequentially over a large set of (x, y) points across the sensor surface, simultaneously providing both surface reconstruction results and data for optical-output-to-force mapping. b) The histograms show empirical distributions of the three position components (P_x, P_y, P_z) along with the corresponding force measurements (F_x, F_y, F_z). c) Regression pipeline. A ResNet-18 backbone (trained from scratch with all layers unfrozen) extracts features from the tactile images, and a lightweight regression head simultaneously predicts the six-dimensional output vector $(P_x, P_y, P_z, F_x, F_y, F_z)$.

range of each sensor to avoid destructive loading. For example, the ViTacTip was limited to ≈3.5 mm ($F_z \leq 0.7$ N), MagicTac to ≈1.2 mm ($F_z \leq 2.85$ N), and GelSight-series to ≈1.0 mm ($F_z \leq 1.83$ N). These limits were established empirically from preliminary trials and consistently adopted in all calibration and evaluation experiments.

In our work, the contact positions based on the robot kinematics $(P_\mathrm{x}, P_\mathrm{y}, P_\mathrm{z})$ and corresponding six-axis force measurements from the F/T sensor $(F_\mathrm{x}, F_\mathrm{y}, F_\mathrm{z})$ serve as ground-truth labels. Each probe instance is paired with a synchronized raw camera image from the VBTS, creating a labeled dataset for supervised learning.

We aggregate exhaustive and randomized samples across the entire sensing area to form a comprehensive dataset. Before training, all labels $(F_\mathrm{x}, F_\mathrm{y}, F_\mathrm{z}, P_\mathrm{x}, P_\mathrm{y}, P_\mathrm{z})$ are min–max normalized to equalize their dynamic ranges. The dataset is partitioned into training, validation, and test sets in a 70:20:10 split. We employ a ResNet-18 backbone for the regression backbone. The network is trained from scratch with randomly initialized weights and optimized with the SGD optimizer for 100 epochs. Model selection is based on the lowest validation error, and final evaluation compares performance metrics across different VBTSs. The resulting ResNet-18 regression model trained from scratch serves as a baseline to compare sensor performance.

Lastly, we evaluate the predictive accuracy for force and contact position across different sensors using the following metrics:

Mean Absolute Error (MAE): Measures the average absolute difference between predictions and ground truth.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\, y_\mathrm{i} - \hat{y}_\mathrm{i}| \tag{1}$$

Coefficient of Determination (R^2): Indicates how well predictions align with actual data trends.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_\mathrm{i} - \hat{y}_\mathrm{i})^2}{\sum_{i=1}^{n} (y_\mathrm{i} - \bar{y})^2} \tag{2}$$

Symmetric Mean Absolute Percentage Error (sMAPE): Quantifies relative error as a percentage by normalizing absolute prediction errors against the average of the true and predicted values

$$\mathrm{sMAPE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_\mathrm{i} - \hat{y}_\mathrm{i}|}{\frac{|y_\mathrm{i}|+|\hat{y}_\mathrm{i}|}{2} + \varepsilon} \times 100\% \tag{3}$$

Here, y_i denotes the ground-truth value, $\hat{y}_\mathrm{i}$ the predicted value, n the total number of samples, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_\mathrm{i}$ the mean of the ground-truth values, and $\varepsilon = 1 \times 10^{-8}$ a small constant to prevent division by zero.

MAE provides an intuitive measurement of average absolute deviations between predictions and ground truth, directly indicating the magnitude of prediction errors. R^2 complements MAE by quantifying the proportion of variance explained by the model. Both MAE and R^2 have been widely used in tactile sensing regression tasks.[11]

sMAPE highlights relative error as a percentage, normalizing prediction errors against signal magnitude. This makes sMAPE particularly effective at highlighting proportional errors and comparing accuracy across different scales. Although less commonly reported in VBTS literature, we introduce it here to complement MAE and R^2 for cross-sensor evaluation.

3.3. Spatial Resolution (SR)

3.3.1. Definitions

The SR of a VBTS denotes the minimum distance between two points on the sensor surface that can be distinctly sensed as separate touch points. This reflects the sensor's capacity to differentiate between closely spaced stimuli. A higher SR score means the sensor can detect finer details of the touched object or surface, making this aspect critical for applications that require detailed tactile feedback, such as dexterous robotic manipulation or surface texture recognition.

3.3.2. Experimental Protocols

We used an Objet500 3D printer to fabricate test samples with linear and cylindrical grating features (in **Figure 3**b), with grating widths and spacings ranging from 0.25 mm to 1.75 mm in 0.05 mm increments. The VBTS was mounted onto the robot's end-effector and used to press each sample 100 times while varying the tool center point (TCP) yaw orientation, producing data, as shown in Figure 3a, right. The entire process is automated with a large number of grating samples placed onto a 3D-printed board with equal spacing (in Figure 3a, the red arrow). This protocol yielded a diverse dataset capturing the sensor's response to samples with different spatial features and orientations.

We used the dataset to train a classifier to distinguish between samples with different grating resolutions. After training, we evaluated accuracy under varying error tolerance thresholds ε, where ε specifies the maximum allowable deviation (in mm) between the predicted and ground-truth resolution levels. Since the gratings' dimensions were fabricated in 0.05 mm increments, all thresholds were chosen as integer multiples of 0.05 mm, ranging from the strictest level (0.05 mm) to more relaxed values (e.g., 0.50–1.00 mm). Rather than reporting a single scalar metric such as cross-entropy loss, we define spatial resolution $\mathrm{SR}(\varepsilon)$ as the curve of the correctly classified samples' accuracy over thresholds ε

$$\mathrm{SR}(\varepsilon) = \frac{1}{N}\sum_{i=1}^{N} \begin{cases} 1, & \text{if} |\hat{r}_i - r_i| \leq \varepsilon, \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\hat{r}_i$ and r_i denote the predicted and ground-truth resolution levels of the i-th sample.

In this way, $\mathrm{SR}(\varepsilon)$ provides a comprehensive characterization, revealing how performance degrades as tolerance becomes stricter (i.e., as $\varepsilon \downarrow$). In addition, it directly shows how VBTS performance varies under different resolution requirements, offering a more intuitive and physically meaningful comparison. This is particularly useful for practical tasks, since not all applications demand the finest resolution, and sensors can be compared under tolerance levels aligned with task needs.

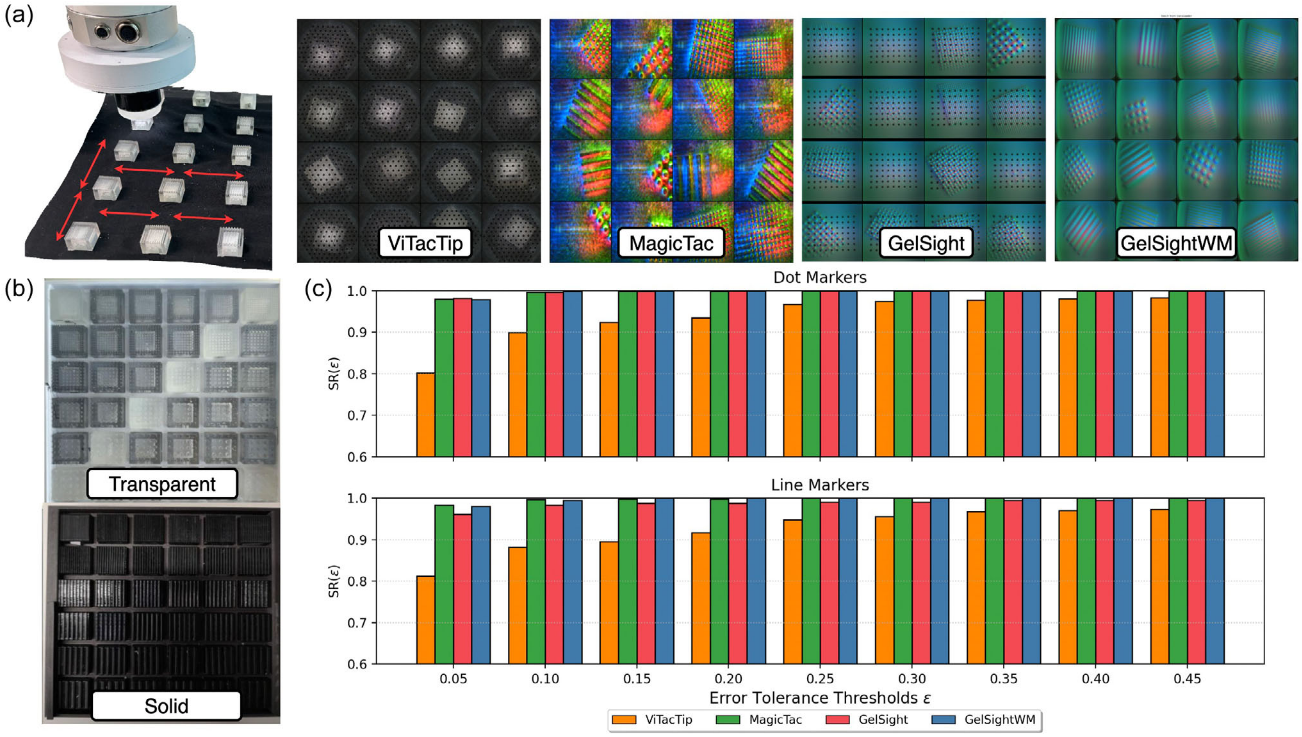

Figure 3. a) Data collection setup using grating boards and raw visual output from selected VBTSs used in the experiment. A black cloth base and sufficient spacing between the grating boards mitigated lighting or positional bias in data collection and prevented transparent VBTSs from exploiting background patterns as cues. b) A total of 72 3D printed line and dot grating boards of both transparent and solid materials were used to conduct the spatial resolution evaluation. c) Classification results of VBTSs showing the effect of different error tolerance thresholds on prediction accuracy (spatial resolution results at higher error tolerance thresholds are omitted, as they led to near-perfect classification performance).

3.4. Mechanical Sensitivity (S)

3.4.1. Definitions

Mechanical sensitivity refers to the compliance of the sensor surface under normal loading, effectively quantifying the indentation response per unit applied force. We define the mechanical sensitivity S as the ratio of indentation depth Δz along the surface normal to the corresponding applied normal force F

$$S = \frac{\Delta z}{F} \tag{5}$$

This definition enables the generation of sensitivity maps across the sensor surface to assess local responsiveness under normal loading. This metric is chosen because it provides a physically interpretable measure of compliance, directly relating normal indentation to applied force. This concept has been highlighted in previous tactile sensor studies in TacTip and GelSight,[3,10,39] where the compliance of the sensor was shown to strongly affect sensing performance, such as force response, durability, and the fidelity of contact features.

Additionally, to quantify the *uniformity* of the sensitivity distribution, we divide the (x, y) plane into uniform bins and compute the mean μ and standard deviation σ of the average sensitivity values across bins. The uniformity score is then defined as

$$U = \frac{1}{1 + \sigma/|\mu|} \in (0, 1] \tag{6}$$

where a larger U indicates greater uniformity. This formulation is mathematically equivalent to a normalized inverse of the coefficient of variation, a standard statistical measure of relative dispersion. Uniform sensitivity distribution is fundamental for reliable VBTS performance, because inconsistent sensitivity can lead to location-dependent bias, distort force estimation, and ultimately reduce the robustness and reliability of tactile perception in tasks like force regression or contact reconstruction.[3,40]

3.4.2. Experimental Protocols

Sensitivity is derived from the calibration dataset (defined in Section 3.2). Each calibration trial provides a contact position (P_x, P_y, P_z) along with ground-truth forces (F_x, F_y, F_z) measured by a six-axis force/torque (F/T) sensor. All probe samples across the sensing surface are aggregated and projected onto the (x, y) plane. The plane is partitioned into spatial bins, and for each bin, we compute the mean sensitivity, which can be visualized as a heatmap. This procedure yields a spatial sensitivity map of the entire sensor surface under normal loading.

3.5. Spatial Robustness (R_{spatial})

3.5.1. Definitions

Spatial robustness refers to how consistently the model performs when the point of contact moves to different locations on the sensor surface. A model with high spatial robustness will give nearly identical outputs regardless of whether the contact is near the center, the edge, or any intermediate point. This is crucial to ensure that the system behaves predictably in real-world scenarios in which the exact contact position cannot be controlled. For each predicted output variable c (i.e., each force or position component), we evaluate spatial robustness along two directions: the sensor surface (radial position) and the indentation range (depth position). The radial position is defined as the Euclidean distance from the sensor center on the tactile surface, while the depth position denotes the indentation magnitude relative to the undeformed sensor surface. Both radial and depth distances are normalized to $[0, 1]$ and uniformly discretized with a bin width of 0.01 ($\approx$100 bins per axis). For each direction, all samples falling into the same bin are grouped together, and we compute both the mean MAE and its standard deviation within that bin.

The per-bin STD values are then smoothed and averaged to obtain a single robustness score for each variable. This score is defined as the combined average of radial and depth variability

$$R_{\text{spatial},c} = \frac{1}{2}\left(\text{mean}\left(\left\{\text{STD}_{b1}^{\text{rad}}\right\}\right) + \text{mean}\left(\left\{\text{STD}_{b2}^{\text{dep}}\right\}\right)\right) \tag{7}$$

where $\text{STD}_{b_1}^{\text{rad}}$ and $\text{STD}_{b_2}^{\text{dep}}$ are the (smoothed) within-bin standard deviations of the errors in each radial and depth bin, respectively, with b_1 and b_2 indexing radial and depth bins. Smaller $R_{\text{spatial},c}$ indicates higher robustness.

3.5.2. Experimental Protocol

We collected a new dataset for each sensor comprising 1,600 indentation points, following the same calibration protocol as in Section 3.2. Model predictions were compared against ground truth to compute MAE and variability (STD).

These errors were then postprocessed to examine their distribution along two dimensions: (i) normalized radial position from the sensor center, reflecting surface uniformity (**Figure** 4b); and (ii) normalized depth position (Figure 4c), reflecting depth uniformity. The resulting trends provide a direct measure of how consistently each sensor maintains accuracy across its spatial domain.

3.6. Lighting Robustness (R_{light})

3.6.1. Definitions

Lighting robustness refers to a model's ability to maintain stable performance under varying illumination conditions, such as bright daylight, dim indoor lighting, or dynamically changing environments. This metric is only applicable to transparent or semitransparent VBTS devices; opaque sensors inherently block external light and are therefore not influenced by lighting variations (their robustness score is nominally deemed as 1, but they are effectively excluded from this evaluation).

We quantify robustness by measuring the model's deviation in prediction accuracy across different lighting conditions and comparing these to a baseline (the original training setting). The metric is defined as follows

$$R_{\text{light}} = \frac{\left|\frac{I_c}{I_o} - 1\right|}{\left|\frac{I_c}{I_o} - 1\right| + \left|\frac{\text{MAE}_c}{\text{MAE}_o} - 1\right|} \tag{8}$$

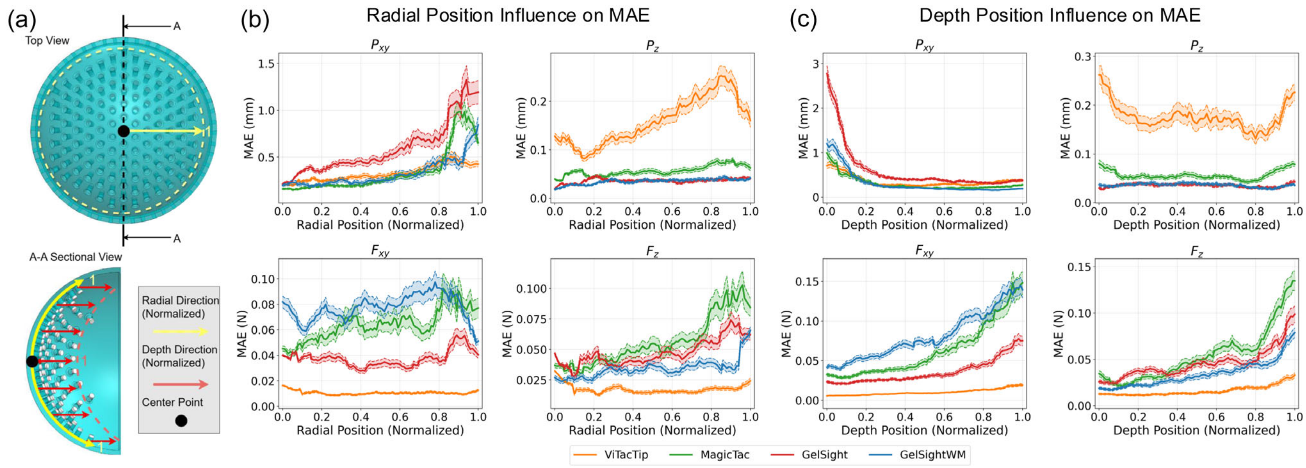

Figure 4. a) Top view and sectional view of the dome-shaped VBTS, defining the radial and depth directions with respect to the sensor center point. b) Radial position influence on MAE: smoothed MAE distributions as a function of normalized radial position for P_{xy}, P_z, F_{xy}, F_z. c) Depth position influence on MAE: smoothed MAE distributions as a function of normalized depth position for the same quantities. Solid lines show the rolling-mean MAE, and shaded bands are $\pm 0.1 \times$ STD. These distributions form the basis for computing the spatial robustness score $R_{\text{spatial},c}$, defined as the average STD of mean within-bin errors along radial and depth directions. The original physical ranges are provided here for reference: ViTacTip (radial: 17 mm; depth: 3.5 mm); MagicTac (radial: 10 mm; depth: 1.0 mm); Gelsight (radial: 12.5 mm; depth: 1.0 mm); GelsightWM (radial: 12.5 mm; depth: 1.0 mm).

where MAE_o is the mean absolute error under the original (training) lighting condition; MAE_c is the mean absolute error under test lighting condition c; I_o is the mean light intensity in the original lighting condition; I_c is the mean light intensity in the test lighting condition c.

A smaller relative change in prediction error as compared to the normalized lighting difference $\left(\frac{I_\text{c}}{I_\text{o}} - 1\right)$indicates minimal performance degradation relative to changes in lighting intensity. Consequently, a higher robustness score R_c (approaching 1) reflects stronger resistance to lighting variations.

3.6.2. Experimental Protocol

To assess the robustness of the lighting, we introduced controlled variations in external illumination during the testing, distinct from the training illumination condition used in calibration in Section 3.2. These included diffuse overhead lighting and point-source illumination, in different combinations and at varying intensities, resulting in four distinct ambient lighting setups (denoted as scenes S1–S4). Specifically, S1 denotes a diffuse ceiling source that produces uniform overhead illumination; S2 employs a lateral wide-beam point source creating broad angular coverage; S3 utilizes a ring-shaped emitter mounted beneath the sensor to provide circumferential lighting around the camera; and S4 applies a narrow-beam directional point source that produces a concentrated spot of illumination. We characterized each lighting condition by the mean gray-scale intensity of the captured images, providing a relative proxy for illumination level across sensors. The baseline model (ResNet-18) trained during calibration was then directly applied to these test sets collected under varied lighting conditions, and its predictive performance under each condition was used to compute the corresponding robustness metrics.

3.7. Repeatability (*Rep*)

3.7.1. Definitions

Repeatability refers to the ability of the sensor to give consistent readings under the same conditions over multiple occasions (i.e., when the same stimulus is applied multiple times). To quantify repeatability, we compute the standard deviation of N repeated measurements at each spatial test point and indentation depth, and then averaged across all test conditions. Specifically, for each output channel c (either positional $P_\text{x}, P_\text{y}, P_\text{z}$ or force $F_\text{x}, F_\text{y}, F_\text{z}$), the standard deviation across N repeated trials is first computed at every spatial point and indentation depth. For each spatial test point $k = 1, \ldots, K$ and indentation step $d = 1, \ldots, D$, N repeated trials are performed, yielding outputs $\hat{c}_{k,d,r}$ $(r = 1, \ldots, N)$. The repeatability score is then defined as

$$\text{Rep}_c = \frac{1}{KD}\sum_{k=1}^{K}\sum_{d=1}^{D}\text{STD}(\hat{c}_{k,d,1}, \ldots, \hat{c}_{k,d,N}) \tag{9}$$

A smaller Rep_c indicates higher repeatability.

3.7.2. Experimental Protocol

A total of $K = 100$ random surface points were selected on the sensor surface, and each point was sampled with 10 repeated trials ($N = 10$). At each point, the sensor was gradually pressed down in steps of 0.1 mm until reaching the predefined maximum depth used during model training, yielding D indentation steps. For every (k, d) combination, $N = 10$ repeated trials were conducted, with the sensor pressed to the target depth and fully retracted before the next trial. This procedure provides both a global repeatability score Rep_c and depth-dependent STD curves for visualization.

4. Case Studies and Evaluation

In this section, we evaluate four representative VBTSs using the evaluation pipeline defined in Section 3, present the results, and provide a comparative analysis of their performance.

4.1. Justification of Sensor Selection

To ensure comprehensive coverage of sensing modalities, we selected four representative VBTSs, as illustrated in Figure 1 (top): 1) ViTacTip: A transparent sensor with embedded markers integrating MDM and MFM.[11] It introduces multimodal perception through its transparent elastomer. This not only enhances deformation estimation but also enables proximity sensing, making it a suitable candidate for evaluating modality fusion. 2) MagicTac: a recent design that integrates IMM, MDM, and MFM simultaneously by using a 3D multilayer grid-based elastomer.[41] This design combines the photometric cues from internal light reflections (IMM), the partial transparency that enables proximity and multimodal perception (MFM), and the continuous markers provided by the embedded grid (MDM). 3) GelSight (standard): An extension of GelSightWM, GelSight incorporates markers on the elastomer surface, thereby combining IMM and MDM. This enables comparison between marker-less and marker-based implementations, highlighting the impact of marker tracking on deformation sensing. 4) GelSightWM: a marker-less, opaque design that is based purely on IMM,[3] representing the most widely adopted class of VBTSs in which dense surface textures are captured through light-intensity variations. This sensor provides a baseline for assessing the performance of marker-less, intensity-only sensing (IMM).

These sensors span the core design spectrum of VBTSs and provide a diverse testbed for validating the generality and applicability of the proposed TacEva framework. For benchmarking, we employ ResNet-18 as a common backbone across all four sensors for the calibration process (Figure 2). Although this offers a consistent baseline, it does not represent state-of-the-art performance, and we would therefore encourage the customization of models tailored to specific sensor characteristics to fully exploit the framework. Moreover, these case studies confirm that the same TacEva protocol, model structure, and metric definitions can be directly transferred across fundamentally different VBTS architectures without modification.

4.2. Experimental Setup

By illustrating TacEva courtesy of case studies using the four VBTSs listed above (which collectively cover the primary sensing modalities), we demonstrate that the framework can accommodate diverse designs and enable meaningful, like-for-like comparison between sensors. All experiments were conducted with a UFactory 850 robot arm, and model training was performed on an NVIDIA RTX 4080 GPU. The intrinsic metrics (physical and optical specifications) of the VBTS are summarized in **Table 1**.

4.3. Force and Localization Calibration Error

Table 2 shows force and localization regression across three important metrics: MAE reflects absolute error in physical units, R^2 quantifies trend fidelity, and sMAPE emphasizes proportional error relative to signal magnitude.

4.3.1. Force regression

ViTacTip attains the lowest MAE in both F_{xy} and F_z, indicating the best absolute force accuracy. Its slightly higher sMAPE arises because its force range is smaller (up to 0.7 N), therefore, identical absolute deviations constitute a larger fraction of signal magnitude. Design-wise, ViTacTip's transparent elastomer with surface markers yields strong, trackable lateral displacements and high mechanical sensitivity per unit load, which reduces absolute error in both F_{xy} and F_z. MagicTac's internal grid supplies partial shear cues, improving F_{xy} over a marker-free design, but depth-dependent refraction and multipath effects increase variance, which is consistent with its higher F_z MAE. GelSight remains balanced with consistently high R^2 (> 0.98), indicating it closely tracks force trends even when absolute errors are not minimal. GelSightWM performs well in F_z (low MAE and sMAPE) but less well in F_{xy}: removing markers produces smoother, more monotonic photometric changes that support normal-force mapping while offering weaker tangential cues.

Table 1. Intrinsic metrics of representative VBTSs.

Metric	ViTacTip	MagicTac	GelSight	GelSightWM
Resolution [MP]	0.9	0.3	8.0	8.0
Thickness [mm]	50.0	15.0	4.25	4.25
FOV [mm^2]	1257	304	266	266
FPS [Hz]	30	60	25	25

Table 2. Performance metrics comparison between sensors.

Sensor	Variable	MAE	R^2	sMAPE
ViTacTip	F_{xy}	**↓0.0102**	0.9617	0.6456
	F_z	**↓0.0164**	0.9775	0.1259
MagicTac	F_{xy}	0.0498	0.9304	0.6081
	F_z	0.0544	0.9675	0.1178
GelSight	F_{xy}	0.0245	**↑0.985**	**↓0.4815**
	F_z	0.0361	0.988	0.0809
GelSightWM	F_{xy}	0.0575	0.893	0.8525
	F_z	0.0269	**↑0.991**	**↓0.0651**
ViTacTip	P_{xy}	0.3514	**↑0.9948**	0.1187
	P_z	0.1781	0.9409	0.1668
MagicTac	P_{xy}	0.2055	0.9925	0.1340
	P_z	0.0516	0.9510	0.1065
GelSight	P_{xy}	0.2479	0.983	0.1485
	P_z	0.0325	0.9735	0.0821
GelSightWM	P_{xy}	**↓0.1445**	0.9927	**↓0.0985**
	P_z	**↓0.0294**	**↑0.9766**	**↓0.0705**

4.3.2. Localization regression

ViTacTip shows high R^2 in lateral positioning alongside larger MAEs in P_{xy} and P_z, because the spherical optical path and dome compliance introduce nonlinear warping and Rim-shadowing, i.e., optical artifacts at the sensor boundary where the dome blocks illumination and reduces edge accuracy, arising from the spherical optical path and dome compliance, causing self-occlusion and uneven lighting.[10] This yields near-correct trends but an offset in absolute coordinates. MagicTac attains a competitive P_{xy} but a weaker P_z, consistent with grid-induced refractive distortions that vary with indentation and confound depth estimation. GelSightWM yields the lowest MAE in both P_{xy} and P_z with high R^2 (>0.97). A likely cause is that its marker-less, reflective surface produces stable photometric gradients, allowing reliable localization without marker texture competing for features. GelSight behaves similarly, but its surface markers can slightly perturb the contact patch appearance, explaining modestly higher MAEs.

The differences mainly come from the marker design. ViTacTip combines a soft elastomer with surface markers: the high compliance produces large, easily visualizable deformations, which explain its low MAE in force regression. However, its limited force range means that proportional accuracy (sMAPE) does not improve to the same extent, so its advantage is confined to absolute error metrics rather than overall scaling. GelSightWM uses the same elastomer as GelSight but without markers, generating a cleaner deformation image; this reduces shear sensitivity in F_{xy} but benefits F_z estimation and yields more reliable localization. MagicTac relies on an internal grid that bends and refracts under shear, giving it better F_{xy} performance than a marker-free design, but the grid itself bends inconsistently under load, introducing optical distortions and reducing stability. These design choices explain the trade-offs we observe in force and position regression.

4.4. Spatial Resolution (SR)

By gradually tightening the spatial tolerance threshold, from broader error allowances to finer ones, we effectively characterize their SR, as shown in Figure 3c. When the spatial tolerance threshold ε is above 5 mm, all sensors exhibit near-perfect accuracy, with negligible differences. However, when the tolerance is tightened to 0.05 mm, performance differences become apparent. At this stricter level, ViTacTip's accuracy drops sharply to 80.6%, while the GelSight series still remains at 99.0% and MagicTac at 98.3%. These results demonstrate that the Gelsight series and Magictac sensor both offer exceptionally high SR, while ViTacTip's performance deteriorates significantly under finer resolution demands.

The differences can be attributed to the mechanical and morphological properties of the elastomer layer. ViTacTip employs a soft, dome-shaped, thick gel (50 mm). This geometry, combined with its high compliance, allows it to conform well to object surfaces, but equally distributes strain and smooths out fine local features, leading to rapid accuracy degradation under stricter tolerance thresholds. By contrast, the GelSight series and MagicTac use stiffer, flatter, and thinner elastomer layers that keep strain confined more locally to the contact region, and maintain sharper indentation traces and thus support sharper SR.

In addition, ViTacTip employs a relatively low-resolution camera combined with a comparatively large sensing field-of-view, which reduces its effective pixel density on the contact surface. When the indenter feature size approaches this imaging limit, the model accuracy deteriorates more rapidly. In contrast, the GelSight series and MagicTac achieve a higher effective sampling density (pixels per mm^2) because their camera resolutions are higher relative to their smaller sensing FOV.

In conclusion, both material properties and morphology have a significant impact on the SR. Soft gels (e.g., PDMS and hydrogel) conform well to the contact surface. However, when the material is too compliant, especially in thick, dome-shaped layers, it can develop global strain that blurs local detail by its internal adhesion and shear. Stiffer, thin-layer designs preserve well-defined indentation patterns, though they require higher loads to conform fully. Therefore, intermediate-modulus transparent elastomers in flat geometries are often considered optimal, as they balance compliance and fidelity in feature capture. Finally, SR is jointly constrained by the imaging quality: a high effective pixel density (camera resolution relative to sensing field-of-view) is essential to sample these features once they are formed on the elastomer surface.

4.5. Mechanical Sensitivity (S)

Figure 5 illustrates the mechanical sensitivity distribution on a logarithmic scale across the four VBTSs. MagicTac, GelSightWM, and GelSight exhibit relatively uniform and low sensitivity levels, reflected by the predominantly blue color scale across their contact regions. In contrast, ViTacTip not only shows higher overall sensitivity values (shifted toward the green–yellow range) but also a pronounced spatial variation: the central region maintains a lower sensitivity of ($S \approx 5$), whereas the edges approach nearly twice this value. This indicates that, for the same applied force, ViTacTip deforms more than the other sensors, especially near the periphery, consequently exhibiting higher effective compliance and stronger boundary effects. For any given F_z, it therefore exhibits a larger displacement P_z, but also a reduced usable force range. However, this higher compliance comes at the cost of reduced uniformity (U), meaning that

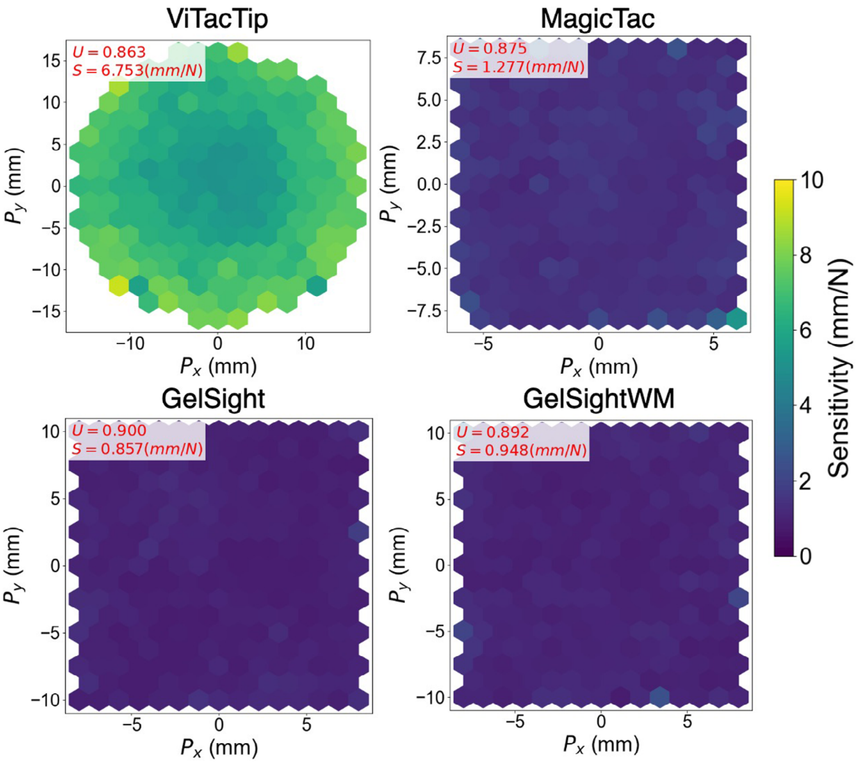

Figure 5. Sensitivity heat map for MagicTac, GelSightWM, GelSight, and ViTacTip, alongside their respective uniformity and average sensitivity values.

predictions of force or displacement may vary considerably depending on contact location. In contrast, the stiffer gels of GelSight and MagicTac maintain higher U values, despite lower S values, suggesting they are better suited for tasks requiring consistent calibration across the surface. In summary, ViTacTip favors sensitivity at the expense of spatial consistency, whereas GelSight-type sensors emphasize stability over the range of deformation.

4.6. Spatial Robustness (R_{spatial})

Figure 4 shows that spatial robustness is clearly both variable and design-dependent, rather than being a uniform edge effect. Edge degradation is strong in the GelSight-series for P_{xy}, F_{xy}, and in all sensors for F_z, while ViTacTip largely resists such trends in the force domain. Depth profiles also diverge: P_{xy} errors drop sharply after shallow contacts, P_z curves are generally flat with occasional U-shapes, and both force channels grow with indentation depth, though much less so for ViTacTip.

A summary of the experimental results P_{xy}, P_z, F_{xy}, F_z for the four representative sensors are outlined below. While MAE-related trends are visualized in Figure 4, The corresponding spatial variability (within-bin STD) is provided separately in **Table 3**.

4.6.1. Lateral Position (P_{xy})

Across all sensors, localization error increases toward the periphery (Figure 4b upper left). ViTacTip attains the lowest overall MAE (<0.5 mm) and variability (STD 0.258 mm), reflecting stable localization afforded by its spherical cavity and uniform imaging geometry. In contrast, GelSight and GelSightWM exhibit a clear error increase beyond when the radial position reaches 80%, consistent with peripheral optical distortion from planar imaging. Normalized-depth analysis (Figure 4c upper left) indicates that the decline in errors arises because shallow contacts create incomplete contact areas that give ambiguous lateral cues, while deeper indentations enlarge and stabilize the contact patch, improving localization.

4.6.2. Normal Position (P_z)

Depth-estimation MAE remains low (< 0.25 mm Figure 4b upper right), increasing only modestly toward the periphery. GelSight-based sensors provide the most stable performance (STD 0.027 mm), because of a thinner, stiffer gel that provides clearer depth cues, while MagicTac shows comparable performance, likewise, due to its thin gel. MagicTac shows comparable performance (STD 0.043 mm), likewise benefiting from its thin gel. In contrast, ViTacTip shows much higher peripheral variance (STD 0.144 mm), likely due to rim shadowing. Both ViTacTip and MagicTac exhibit a mild U-shaped trend with depth (Figure 4c upper right). Errors are greater when contact is shallow (insufficient compression for reliable photometric cues) and, equally, when indentations are very deep (optical distortion near the gel boundary).

Table 3. Overall spatial robustness (mean within-bin STD across radial and depth conditions).

Metric	ViTacTip	MagicTac	GelSight	GelSightWM
P_{xy}	↓**0.258**	0.266	0.484	0.300
P_z	0.144	0.043	↓**0.028**	↓**0.027**
F_{xy}	↓**0.009**	0.069	0.032	0.061
F_z	↓**0.025**	0.125	0.109	0.147

Note: Position metrics (P_{xy}, P_z) are in mm; force metrics (F_{xy}, F_z) are in N. Bold entries with ↓ indicate the best (lowest) spatial variability within each row. Lower values indicate higher spatial robustness.

4.6.3. Lateral Shear Force (F_{xy})

This performance metric depends on marker-displacement visibility. ViTacTip outperforms the other sensors, maintaining low error and variance across the surface (STD 0.009 N). In contrast, GelSightWM (STD 0.061N) suffers from the absence of markers, while MagicTac's internal grid introduces depth-dependent refraction that destabilizes feature tracking. Comparison analysis against normalized depth (Figure 4c lower left) further reveals that MAE increases monotonically with indentation depth for all GelSight-based sensors, indicating that shear-force estimation becomes progressively more difficult as marker deformations intensify.

4.6.4. Normal Force (F_z)

ViTacTip again achieves the lowest normal-force MAE (<0.02 N Figure 4b lower right) and the lowest spatial variability (STD 0.025 N), maintaining stable performance across both spatial dimensions. The other three sensors show substantial error growth toward the edges as well as with increasing depth, most likely due to boundary stiffness gradients in planar elastomers and reduced uniformity of vertical marker motion near the edges. This highlights a fundamental limitation of flat-gel architectures for normal force estimation.

In summary, position-localization errors (P_{xy}, P_z) are driven primarily by optical and geometric distortions rather than marker density. In contrast, the accuracy of the force estimation (F_{xy}, F_z) depends critically on the visibility of the marker and the uniformity of the displacement of the marker, with the geometry of the sensor and the stiffness of the edges also playing a substantial role. ViTacTip's spherical geometry and dense, uniform marker pattern deliver high accuracy and stable performance in force estimation, whereas planar GelSight-based sensors exhibit spatial degradation, particularly toward the periphery and under deeper indentations.

4.7. Lighting Robustness (R_{light})

For this metric, we restrict evaluation to MFM-based sensors, which provide multimodal sensing but are susceptible to external light sources. We created new test data by setting up four lighting scenes (**Figure 6**a) that are different from the illumination condition of the dataset used in calibration. As shown in the raw visual output under the different lighting scenes, the ViTacTip

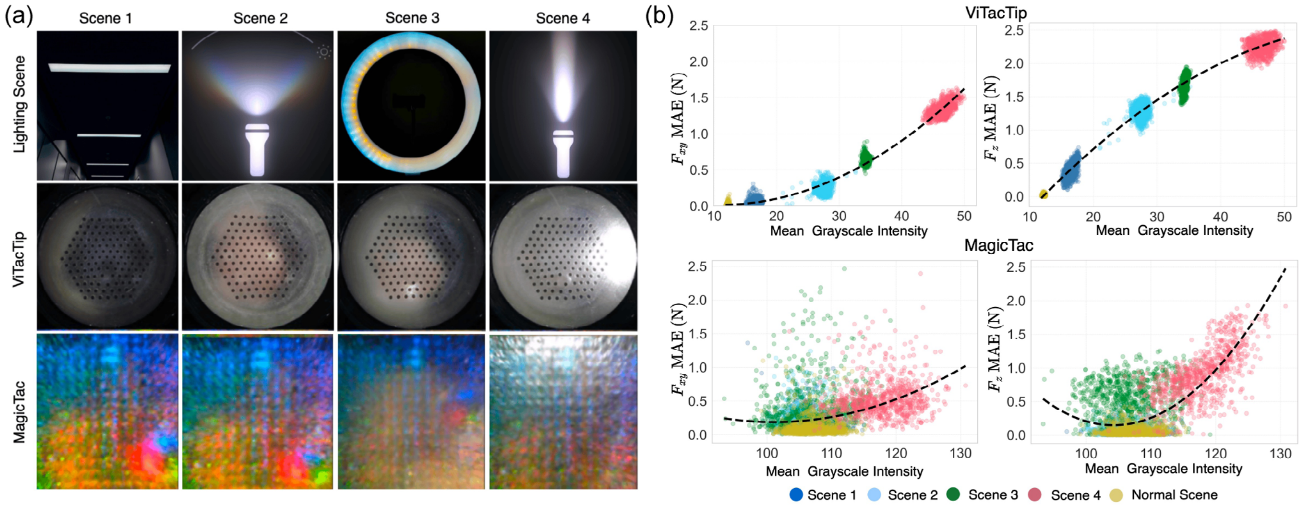

Figure 6. a) Lighting scenes versus visual output of ViTacTip and MagicTac. Four different lighting scenes, notably dim room lighting, diffuse flash lighting, streamer ring lighting, and streamed flash lighting, were tested to evaluate their effect on the sensor's visual output. Sample frames are shown from ViTacTip (top) and MagicTac (bottom) under four illumination conditions. b) Force-regression MAE plotted against mean grayscale intensity: ViTacTip shows increasing error with brightness, while MagicTac stays comparatively stable.

sensor is highly susceptible to changes in lighting intensity due to its material properties and geometry.

In Figure 6b, the maximum light intensity in the experiments was up to ≈7 times greater than that experienced during training conditions. In contrast, MagicTac exhibits greater resistance to external light noise due to the refraction effects caused by its internal grid structure. As a result, diffuse light sources (scenes 1–3) had minimal impact on its average light intensity, with only point light sources causing noticeable fluctuations. This suggests that MagicTac effectively mitigates the influence of occlusions on illumination. Although MagicTac's physical design offers a degree of light shielding (resulting in small intensity deviations), the prediction error variance increases markedly under the same lighting condition, resulting in poorer model robustness than ViTacTip despite stable raw intensities. Under point light conditions (scene 4), ViTacTip's MAE increases by a factor of about 10 relative to its performance under normal lighting conditions, even though the light intensity is only about seven times higher. In contrast, with MagicTac, the intensity-induced error is much lower, but MAE values have high variance, which likely arises from its lattice-embedded markers. Variations in lighting make it harder to align the captured grid with the training reference, reducing marker distinctiveness and impairing tracking. Furthermore, MagicTac leverages photometric stereo principles by utilizing red, green, and blue light to enhance z-axis depth and force regression. When external illumination interferes with the internally generated light, errors inherent to the photometric stereo method may also arise.

In summary, ViTacTip demonstrates higher robustness for tangential force and position estimation, while MagicTac shows superior robustness in normal force estimation. For normal positions, the two sensors perform comparably, with ViTacTip showing a slight overall advantage.

4.8. Repeatability (*Rep*)

We used the ResNet-based model calibrated for force and contact-localization regression as predictors. **Figure 7**a shows predictions from repeated trials across different indentation planes, and Figure 7b illustrates the mean standard deviation (STD) curves as a function of indentation depth. These depth-dependent plots illustrate how repeatability varies, while the aggregated Rep_c values in Table 5 provide a global comparison.

For GelSightWM, repeatability under shear-dominant stimuli is not reported because its marker-less, intensity-mapping design provides weak observability of tangential components; accordingly, we focus on sensors with explicit displacement cues for shear-dominant stimuli.

For the predicted force variables (F_{xy}, F_z), ViTacTip consistently demonstrates the most stable force predictions across various depths (lowest STDs; F_{xy}: 0.006, F_z: 0.007). The next best is GelSight (F_{xy}: 0.025, F_z: 0.026) followed by MagicTac (F_{xy}: 0.041, F_z: 0.040). Primarily attributable to ViTacTip's soft dome shape, it produces large, smooth deformations under load, which are easier for the model to regress consistently. MagicTac varies more because even small differences in repetitive presses can easily change the internal grid's structure, which the model interprets as force changes.

In terms of predicted positional variables (P_{xy}, P_z), ViTacTip again demonstrates the best repeatability (P_{xy}: 0.166), though it shows higher variability at very shallow contacts due to unstable skin deformation under minimal loading. GelSight achieves the most stable depth localization (P_z: 0.015), but performs poorly in lateral localization (P_{xy}: 0.278). This suggests that GelSight's opaque elastomer geometry, rather than its surface markers, governs its positional repeatability, indeed, the dense marker layer may add noise rather than provide useful cues for

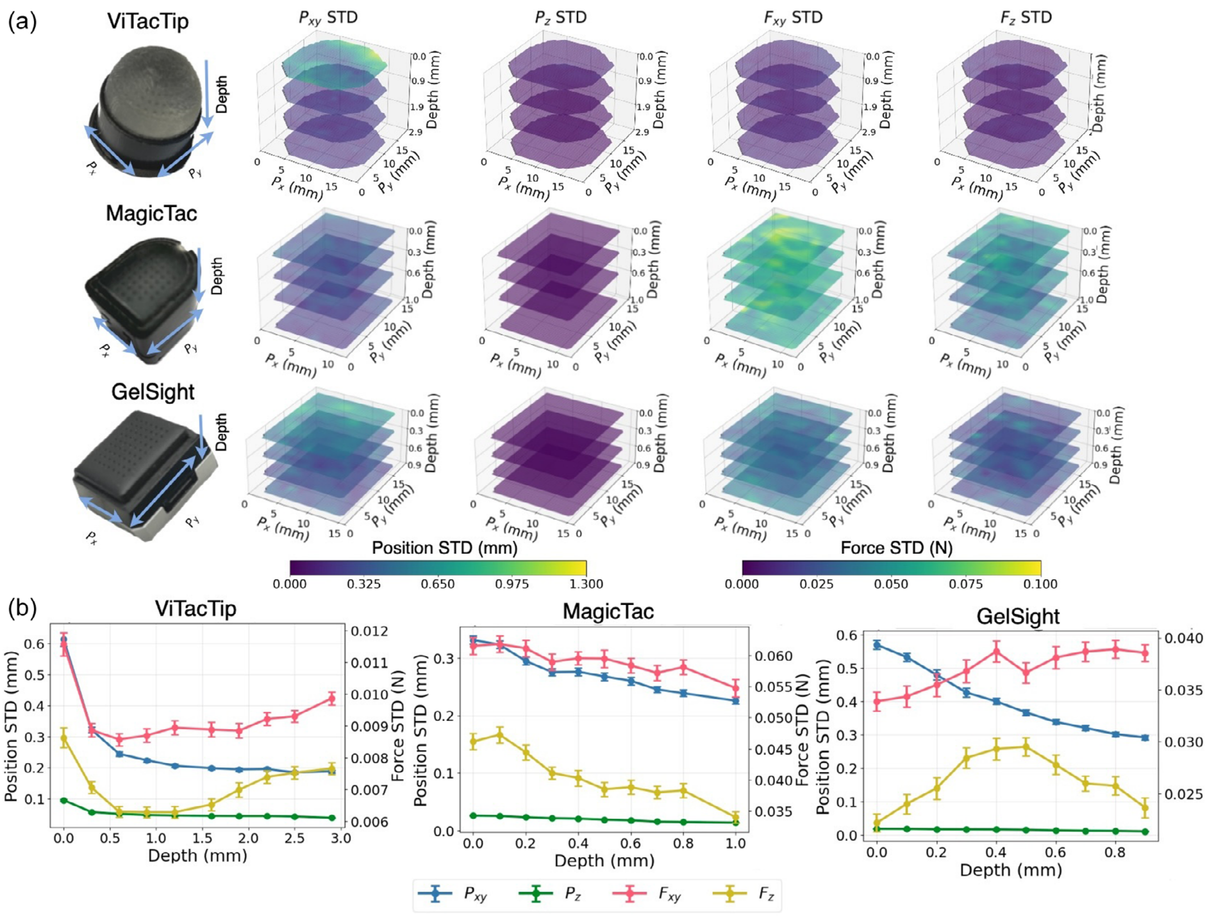

Figure 7. Repeatability analysis for ViTacTip, MagicTac, and Gelsight. a) Standard deviation (STD) of predicted force components (F_{xy}, F_z) and positional components (P_{xy}, P_z) across multiple trials at various depths of VBTS. b) Line plots of repeatability (STD) of the predicted outputs as a function of indentation depth. Positional STDs (P_{xy}, P_z, left y-axis, in mm) and force STDs (F_{xy}, F_z, right y-axis, in N) are shown for each sensor.

localization. MagicTac shows intermediate repeatability (P_{xy}: 0.188, P_z: 0.020), performing better laterally than GelSight, but worse axially.

Overall, ViTacTip performs best on repeatability, benefiting from its compliant dome and optical transparency. Variability is only notable during the earliest shallow contacts, when the visible cues are too subtle. GelSight is particularly stable in depth localization, but its lateral repeatability is limited. The opaque elastomer and dense marker distribution appear to constrain reliable P_{xy} tracking. While MagicTac provides intermediate positional repeatability, its force estimates remain the most variable, as the internal grid marker design is prone to instability under repeated loading. Notably, ViTacTip was operated over a wider indentation span (3.5 mm) but a narrower force range (0.7 N), whereas the others covered shallower depths (1 mm) with higher forces (beyond 2 N). This indicates that its soft and compliant dome converts small forces into large, smooth deformations, producing highly stable cues under low loading. By contrast, stiffer designs deform less but tolerate larger forces, so their repeatability is reduced even though they can cope with stronger contacts. This suggests that repeatability is inherently linked to material compliance, and therefore, that sensors designed for stable low-force interaction should favor softer, thicker elastomers, while stiffer gels would be better suited to tasks requiring force endurance.

4.9. Sensor-Level Evaluation and Decision Guide

Figure 8 compares the four VBTSs on three categories: intrinsic properties, standard performance, and robustness. The radar plots' numerical values are taken from the corresponding quantitative tables and figures. Specifically, for the radar plot (a), values are extracted from Table 1. For radar plot (b), the MAE and related performance metrics are taken from **Table** 2, in which the SR corresponds to Figure 3d, with a tolerance set to 0.05 mm, and sensitivity values are given in Figure 5. For radar plot (c), lighting robustness values are taken from **Table 4**, while spatial robustness and repeatability are taken from **Table** 3 and **5**, respectively. As the original tables report F_{xy} and F_z separately, their mean is computed to yield the "Force" axis in the radar chart; similarly, P_{xy} and P_z are averaged out to obtain the "Position" axis.

Intrinsic properties are related to the hardware specifications of the sensors. These metrics reflect the fundamental design trade-offs imposed by the sensor's hardware and optical

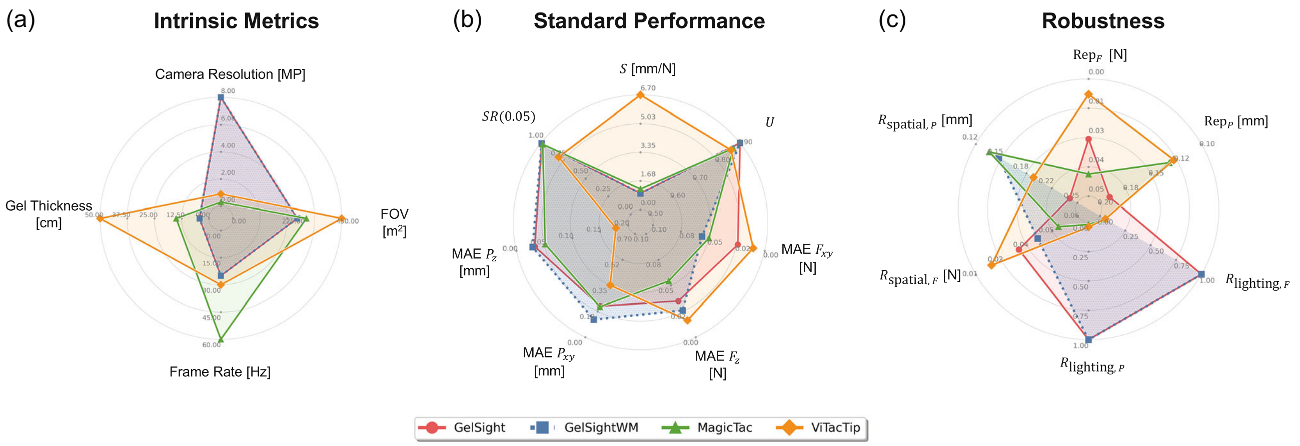

Figure 8. Radar plots comparing the four VBTSs on three evaluation themes: a) Intrinsic properties (camera resolution, field of view, frame rate, gel thickness, and sensitivity). b) Standard performance (spatial resolution and MAEs in force and position, and sensitivity). c) Robustness (including repeatability, spatial robustness, and light robustness in both force and position metrics). Larger areas correspond to stronger performance across the evaluated metrics.

Table 4. Light robustness comparison between ViTacTip (left) and MagicTac (right).

Scene	F_{xy} [%]	F_z [%]	P_{xy} [%]	P_z [%]
S1	11.5/1.8	1.2/35.1	2.0/2.7	2.0/4.7
S2	17.6/1.2	1.4/2.8	3.9/1.1	3.3/4.2
S3	4.4/0.1	1.5/0.0	6.5/0.0	3.2/0.0
S4	3.4/2.3	1.7/0.6	7.3/1.0	3.6/1.4
Mean	↑**9.3**/1.4	1.5/↑**9.6**	↑**5.0**/1.2	↑**3.0**/2.6

Table 5. Repeatability results for position and force variables.

Sensor	P_{xy}	P_z	F_{xy}	F_z
ViTacTip	↓**0.166**	0.049	↓**0.006**	↓**0.007**
GelSight	0.278	↓**0.015**	0.025	0.026
MagicTac	0.188	0.020	0.041	0.040

pathways. Standard performance characterizes the sensor's accuracy and resolution under standard conditions. This includes regression performance (e.g., force/pose estimation error), compliance of the elastomer, and SR for detecting fine contact patterns. These indicators capture the sensor's "nominal" capability when operating without external perturbations. Robustness is a measure of how consistently the sensors maintain their performance under varying conditions. The table includes repeatability and spatial robustness, quantified using the standard deviation of model predictions to reflect how stable the output remains across trials and positions. Lighting robustness is also reported as a dimensionless parameter, capturing the sensitivity of each sensor to illumination changes. These results reveal distinct trade-offs across categories, indicating different application niches for each sensor type.

ViTacTip achieves relatively strong force accuracy and the best repeatability, reflecting highly stable behavior across different stages. However, its optical path makes it sensitive to external illumination, and its depth estimation (P_z) and SR are weaker. Materially, it has the softest gel, the greatest thickness, and the largest field of view. These features make it well-suited to tasks emphasizing force estimation on soft or deep contacts, where fine texture discrimination and sub-millimeter localization are less critical.

MagicTac excels in contact-localization accuracy and uniformity, with SR only slightly behind the GelSight series, while also offering the highest frame rate. In contrast, its force estimation is generally weaker, and its illumination robustness is poor, likely a consequence of the internal grid producing refractive artifacts under varying light conditions. This profile favors fast, planar localization in scenes where force accuracy demands are modest and illumination can be controlled.

GelSight offers the highest camera resolution and the most uniform gel sensitivity, which is beneficial to fine-grained contact localization. These advantages come with trade-offs: the gel is comparatively stiff (lower sensitivity), and the camera's frame rate is modest. Edge-area distortion is also evident in the repeatability panel. Nevertheless, its force response is relatively uniform across the field, yielding a well-rounded overall profile.

GelSightWM delivers stable performance in normal force and depth estimation (F_z/P_z) but remains weaker in shear force (F_{xy}). Its planar contact localization (P_{xy}) is competitive among the four; the absence of surface markers likely encourages the model to rely more directly on contact-induced cues rather than visual texture, which benefits contact detection in the plane.

For deep or soft contacts, where normal-force sensitivity is central, ViTacTip is the obvious choice despite its light sensitivity. Shallow or firm contacts benefit from GelSight or MagicTac, with GelSight preferred for depth (P_z) and MagicTac for planar localization (P_{xy}) when speed matters. Applications that stress accurate shear force should still prioritize ViTacTip, but when robustness to illumination is paramount, GelSight/GelSightWM are safer defaults. In scenarios where shear is not required, GelSightWM provides a practical and sufficiently stable solution.

4.10. Application-Level Interpretation of TacEva Metrics

Beyond comparing individual sensors, the radar plots also help interpret how TacEva's metrics relate to real-world task requirements. From an application perspective, VBTS-related tasks can be broadly categorized into three classes: geometry-focused, force-focused, and robustness-critical tasks.[1,4,38,42,43]

Geometry-focused tasks (e.g., edge following, in-hand pose control, texture refinement, contact reconstruction) primarily require fine-grained contact geometry and accurate planar

Table A1. Reported evaluation metrics across VBTS works, grouped by sensing mechanism proposed in ref. [8].

Work/sensor	(a) Intrinsic Metrics				(b) Standard Performance			(c) Robustness		
	Res.	FOV	Gel	FPS	Calib.	*SR*	*S*	*Rep*	R_{spatial}	R_{light}
IMM										
DIGIT[29]	✓	✓	✗	✓	✗	✓	✗	✗	✗	✗
DTact[49]	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓
OmniTact[30]	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗
Minsight[50]	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗
ThinTact[32]	✓	✓	✓	✓	✗	✓	▲	✗	▲	✗
InSight[51]	✓	✓	✓	✓	✓	▲	✓	✗	✓	✗
GelTip[34]	✗	✗	✓	✗	▲	✓	✗	✗	✗	✗
MDM										
GelForce[28]	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗
ChromaTouch[52]	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗
Tactip[10]	✓	✓	✓	✓	✓	▲	✓	✗	✓	✗
Deltac[53]	✓	✓	✓	✗	✓	✗	✓	✗	▲	✗
Tac3D[54]	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗
IMM + MDM										
GelSight[3]	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗
GelSlim[39]	✓	✓	✗	✓	✓	✗	✗	✓	✓	✓
UV-Tac[55]	✓	✓	✓	✓	✓	▲	✓	✗	▲	✓
DenseTac 2.0[31]	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
IMM + MFM										
VisTac[56]	✓	✓	✓	✓	▲	▲	✗	✗	✗	✗
Finger-STS[57]	✓	✓	✓	✗	✗	✓	✗	✗	✗	✗
TIRgel[58]	✓	✓	✗	✓	✗	✓	✗	✗	✗	✓
HiVTac[59]	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗
MDM + MFM										
ViTacTip[11]	✓	✓	✗	✗	✓	✓	✓	✗	✗	✓
FingerVision[12]	✓	✓	✓	✓	✗	✗	▲	✗	▲	▲
SpecTac[60]	✗	✓	✓	✓	✓	✗	✗	✗	✗	✗
VPTS[61]	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
IMM + MDM + MFM										
MagicTac[41,62]	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓
F-Touch[9]	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓
Evaluation/review works										
PBR-Design[36]	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓
VT-Review[63]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Eval[37]	✗	✗	✗	✓	✗	✓	✓	✗	✗	✗
Sparsh[38]	✗	✗	✗	✓	✗	✓	✗	✗	▲	▲
TacEva (Ours)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Symbols: ✓ = reported/available; ✗ = not reported; ▲ = partially reported or indirectly evaluated.

Table A2. VBTS sensors and their evaluations based on application level interpretation.

Sensor/paper	Geometry-focused	Force-focused	Robustness-critical
IMM			
DIGIT[29]	Pose control; contact & pose tracking	–	–
DTact[49]	Contact reconstruction; pose estimation; object recognition	–	Illumination robustness
OmniTact[30]	Connector insertion; contact angle estimation	–	–
Minsight[50]	Tactile servoing; lump detection	Force estimation	–
ThinTact[32]	Delicate grasping; insertion; sliding pose manipulation	Gentle force regulation	–
InSight[51]	Shape; orientation; posture sensing	Multi-contact force mapping	–
GelTip[34]	Contact localization	–	–
MDM			
GelForce[64]	–	Traction-field estimation	–
ChromaTouch[52]	Curvature estimation (3D displacement field)	–	–
TacTip[10]	In-hand rolling and reorientation; edge following contact localization; JND discrimination	–	–
DelTact[54]	Contact reconstruction	Force mapping	–
Tac3D[53]	Spatial resolution; contact reconstruction	Force estimation; friction-coefficient estimation; real-time force mapping	–
IMM + MDM			
GelSight[3]	Contact reconstruction; texture recognization; USB insertion	Force, stiffness estimation; slip detection	–
GelSlim[33,39,65]	Pose & shape reconstruction; insertion	Force/torque estimation; force-controlled manipulation; slip detection	Durability (>3000 grasps)
DenseTac 2.0[31,44]	Dense contact reconstruction	Force/torque estimation	–
UV-Tac[55]	Contact localization; alignment	Normal force mapping; shear force estimation	UV/white-light switching
IMM + MFM			
VisTac[56]	Contact reconstruction and localization; pose estimation; insertion	–	–
Finger-STS[57]	Object/texture recognition	Dense slip detection; bead-maze tasks	–
TIRgel[58]	Object classification	–	Ambient-brightness resistance
HiVTac[59]	Pose estimation; deformation analysis	Force estimation; grasping;	–
MDM + MFM			
ViTacTip[11]	Grating classification; pose estimation; localization	Force estimation	Lighting robustness (GAN-based)
FingerVision[12,66–68]	Surface reconstruction; deformation field	Force estimation; slip/vibration detection	Grasp-stability tests (shaking)
SpecTac[60]	3D triangulation; feature matching	Force estimation	SIFT feature detection
VPTS[61]	Contact reconstruction; proximity exploration	–	Vision–proximity–tactile fusion
IMM + MDM + MFM			
MagicTac[41,62]	Grating classification; pose estimation; contact localization	Force estimation	Lighting robustness; manufacture error; wear and tear
F-Touch[9]	Object/texture recognition	Force/torque estimation	–
Evaluation and reviews			
PBR-Design[36]	Contact reconstruction (embossed text, grasping); VBTS design	–	Design robustness (optical design stability)
VT-Review[63]	3D visuo-tactile contact reconstruction	Force estimation	Robustness across textures & marker layouts
Eval[37]	Edge detection; contact reconstruction	Minimum detectable forces; force sensitivity mapping; slip detection; frequency tests	Temperature & material dependency; real-world grasping

Table A2. Continued.

Sensor/paper	Geometry-focused	Force-focused	Robustness-critical
Sparsh[38]	Pose tracking; texture recognition	Force estimation and mapping; slip-accumulation tracking; bead-maze tasks; grasping stability	–
TacEva (Ours)	Spatial resolution; contact localization	Force estimation; force sensitivity mapping	Repeatability; spatial & lighting robustness

localization.[3,12,15,44] In TacEva, these correspond to high SR, low localization error, and strong spatial robustness of position estimates. A developer facing a geometry-dominated task should therefore primarily check the "Position", "SR", and "Spatial robustness" axes of the radar plots.

Force-focused tasks (e.g., compliant grasping, delicate object manipulation, exploration with force limits) emphasize accurate and stable force estimation,[7,9,45] captured by force sensitivity, low MAE in F regression, and the repeatability scores. For such tasks, the relevant axes in Figure 8 are "Force", "Sensitivity", and "Repeatability". In tasks that also involve rapid contact changes, such as slip detection, temporal bandwidth,[38,39] also becomes important. In TacEva, this is reflected by the intrinsic FPS of each sensor.

Robustness-critical tasks (e.g., mobile manipulation, tool use in clutter, or any operation under variable illumination) typically arise in unstructured or changing environments where neither lighting nor precise contact location can be controlled.[46–48] TacEva captures these requirements mainly through robustness metrics, which quantify how well a sensor maintains its performance under domain shifts and repetitive use.

These three task classes provide a simple way to read the radar plots: each class highlights a different subset of metric categories. TacEva thus serves as a decision tool: by identifying the task's major requirement and checking the corresponding metrics in Figure 8, developers can select the sensor whose intrinsic properties, nominal performance, and robustness profile best match the operational constraints. To make this mapping explicit, Table A2 groups representative VBTS sensors and evaluation works by sensing mechanism, and for each work indicates which task classes (geometry, force, robustness) they primarily target. Together with Figure 8, these summaries provide a concrete guide for matching TacEva metrics, existing VBTS designs, and application-level task requirements.

5. Conclusions and Future Work

In this work, we introduced TacEva, a systematic evaluation framework for VBTSs. TacEva defines a comprehensive set of standardized performance metrics, provides structured and reproducible evaluation pipelines, and demonstrates their application across multiple sensors with distinct sensing mechanisms. TacEva delivers the first unified framework for consistent and quantitative assessment of VBTS designs, enabling direct comparison and performance-guided insights into sensor optimization.

Looking ahead, the next step is to extend TacEva into a dedicated, community-maintained framework that catalogs a broad range of VBTSs with their core attributes, while also hosting benchmark datasets and standardized protocols for representative tasks. Furthermore, the scope of evaluation should be expanded to encompass characterization of VBTS performance under real robotic manipulation conditions, and naturally connect TacEva to task-level and system-level performance metrics. Collectively, these developments would establish TacEva as a comprehensive platform that integrates shared resources with application-oriented evaluation protocols, thereby enabling objective comparison, informed sensor selection, and performance-driven innovation for the next generation of VBTSs.

Appendix A1: Summary of Reported Evaluation Metrics across VBTS

Table A1 summarizes whether VBTS designs report intrinsic, standard-performance, and robustness metrics. Sensors are grouped by sensing mechanism (IMM, MDM, MFM, and hybrid variants). The metric abbreviations are: Resolution (Res), Field of View (FOV), Gel Thickness (Gel), Frame Rate (FPS), Force and Localization Calibration Error (Calib), Spatial Resolution (SR), Mechanical Sensitivity (SS), Repeatability (Rep), Spatial Robustness (R_{spatial}), and Lighting Robustness (R_{light}).

The table highlights the inconsistency of evaluation practices across prior works, including the lack of quantitative calibration, repeatability, and spatial robustness, etc, in most studies.

Appendix A2: Task Classification of Existing VBTS Evaluation Works

Table A2 categorizes existing evaluation works according to the task perspectives commonly used in VBTS research. Tasks are grouped into three classes, geometry-focused, force-focused, and robustness-critical, reflecting the application-level interpretation introduced in the main text. This overview provides a complementary perspective to the metric summary in Appendix A by illustrating the distribution of task types addressed in prior VBTS studies.

Acknowledgements

Q.C., S.O., and W.F. contributed equally to this work. The authors would like to acknowledge the Department of Bioengineering, Imperial College London, for funding this project. The authors would like to thank Mish Toszeghi (Queen Mary University of London) for helpful comments and proofreading assistance.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in [Tactile-Sensor-Website] at [https://zeyuanx0x.github.io/Tactile-Sensor-Website/], reference number [0].

Keywords

multimodal sensing, performance evaluation, vision-based tactile sensors

Received: September 22, 2025
Revised: December 4, 2025
Published online:

---

[1] A. Billard, D. Kragic, Trends and Challenges in Robot Manipulation, *Science* **2019**, *364*, eaat8414.

[2] Z. Zhao, W. Li, Y. Li, T. Liu, B. Li, M. Wang, K. Du, H. Liu, Y. Zhu, Q. Wang, K. Althoefer, S. Zhu, Embedding High-Resolution Touch Across Robotic Hands Enables Adaptive Human-Like Grasping, *Nat. Mach. Intell.* **2025**, *7*, 889.

[3] W. Yuan, S. Dong, E. H. Adelson, Gelsight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force, *Sensors* **2017**, *17*, 2762.

[4] S. Zhang, Z. Chen, Y. Gao, W. Wan, J. Shan, H. Xue, F. Sun, Y. Yang, B. Fang, Hardware Technology of Vision-Based Tactile Sensor: A Review, *IEEE Sens. J.* **2022**, *22*, 21410.

[5] D. Zhang, W. Fan, J. Lin, H. Li, Q. Cong, W. Liu, N. F. Lepora, S. Luo, Design and Benchmarking of a Multi-Modality Sensor for Robotic Manipulation with Gan-Based Cross-Modality Interpretation, *IEEE Trans. Robot.* **2025**.

[6] B. He, Q. Miao, Y. Zhou, Z. Wang, G. Li, S. Xu, Review of Bioinspired Vision-Tactile Fusion Perception (vtfp): From Humans to Humanoids, *IEEE Trans. Med. Robot. Bionics* **2022**, *4*, 875.

[7] B. Fang, J. Zhao, N. Liu, Y. Sun, S. Zhang, F. Sun, J. Shan, Y. Yang, Force Measurement Technology of Vision-Based Tactile Sensor, *Adv. Intell. Syst.* **2024**, 2400290.

[8] W. Fan, H. Li, D. Zhang, Crystaltac: 3D-Printed Vision-Based Tactile Sensor Family Through Rapid Monolithic Manufacturing Technique, *arXiv preprint arXiv:2408.00638* **2024**.

[9] W. Li, A. Alomainy, I. Vitanov, Y. Noh, P. Qi, K. Althoefer, F-Touch Sensor: Concurrent Geometry Perception and Multi-Axis Force Measurement *IEEE Sens. J.* **2021**, *21*, 4300.

[10] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, N. F. Lepora, The Tactip Family: Soft Optical Tactile Sensors With 3D-Printed Biomimetic Morphologies, *Soft Robot.* **2018**, *5*, 216.

[11] W. Fan, H. Li, W. Si, S. Luo, N. Lepora, D. Zhang, Vitactip: Design and Verification of a Novel Biomimetic Physical Vision-Tactile Fusion Sensor, *arXiv preprint arXiv: 2402.00199* **2024**.

[12] A. Yamaguchi, Fingervision with Whiskers: Light Touch Detection with Vision-Based Tactile Sensors, in *2021 Fifth IEEE Inter. Conf. on Robotic Computing (IRC)*, IEEE **2021**, pp. 56–64.

[13] Z. Song, R. Yu, X. Zhang, K. W. Sou, S. Mu, D. Peng, X.-P. Zhang, W. Ding, Satac: A Thermoluminescence Enabled Tactile Sensor for Concurrent Perception of Temperature, Pressure, and Shear, *arXiv preprint arXiv:2402.00585* **2024**.

[14] Z. Zhao, W. Li, Y. Li, T. Liu, B. Li, M. Wang, K. Du, H. Liu, Y. Zhu, Q. Wang, K. Althoefer, S. Zhu, Embedding High-Resolution Touch Across Robotic Hands Enables Adaptive Human-Like Grasping, *Nat. Mach. Intell.* **2025**, *7*, 889.

[15] N. F. Lepora, Soft Biomimetic Optical Tactile Sensing with the Tactip: A Review, *IEEE Sens. J.* **2021**, *21*, 21131.

[16] T. P. Tomo, M. Regoli, A. Schmitz, L. Natale, H. Kristanto, S. Somlor, L. Jamone, G. Metta, S. Sugano, A New Silicone Structure for Uskin–A Soft, Distributed, Digital 3-Axis Skin Sensor and its Integration on the Humanoid Robot Icub, *IEEE Robot. Autom. Lett.* **2018**, *3*, 2584.

[17] A. Alfadhel, M. A. Khan, S. Cardoso de Freitas, J. Kosel, Magnetic Tactile Sensor for Braille Reading, *IEEE Sens. J.* **2016**, *16*, 8700.

[18] A. Schmitz, M. Maggiali, L. Natale, B. Bonino, G. Metta, A Tactile Sensor for the Fingertips of the Humanoid Robot Icub, in *2010 IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems*, IEEE **2010**, pp. 2212–2217.

[19] A. L. Trejos, J. Jayender, M. Perri, M. D. Naish, R. V. Patel, R. Malthaner, Robot-Assisted Tactile Sensing for Minimally Invasive Tumor Localization, *Int. J. Robot. Res.* **2009**, *28*, 1118.

[20] S. Stassi, V. Cauda, G. Canavese, C. F. Pirri, Flexible Tactile Sensing Based on Piezoresistive Composites: A Review, *Sensors* **2014**, *14*, 5296.

[21] H. Pei, J. Jing, Y. Chen, J. Guo, N. Chen, 3D Printing of Pvdf-Based Piezoelectric Nanogenerator from Programmable Metamaterial Design: Promising Strategy for Flexible Electronic Skin, *Nano Energy* **2023**, *109*, 108303.

[22] J. A. Fishel, G. E. Loeb, Sensing Tactile Microvibrations with the Biotac–Comparison with Human Sensitivity, in *2012 4th IEEE RAS & EMBS international conference on biomedical robotics and biomechatronics (BioRob)*, IEEE **2012**, pp. 1122–1127.

[23] G. Kim, D. Hwang, Barotac: Barometric Three-Axis Tactile Sensor with Slip Detection Capability, *Sensors* **2022**, *23*, 428.

[24] S. Somlor, R. Hartanto, A. Schmitz, S. Sugano, A Novel Tri-Axial Capacitive-Type Skin Sensor, *Adv. Robot.* **2015**, *29*, 1375.

[25] P. Yu, W. Liu, C. Gu, X. Cheng, X. Fu, Flexible Piezoelectric Tactile Sensor Array for Dynamic Three-Axis Force Measurement, *Sensors* **2016**, *16*, 819.

[26] V. Pattabiraman, Z. Huang, D. Panozzo, D. Zorin, L. Pinto, R. Bhirangi, eflesh: Highly customizable magnetic touch sensing using cut-cell microstructures **2025**. [Online]. Available: https://arxiv.org/abs/2506.09994.

[27] R. J. Kirschner, A. Kurdas, K. Karacan, P. Junge, S. A. Baradaran Birjandi, N. Mansfeld, S. Abdolshah, S. Haddadin, Towards a Reference Framework for Tactile Robot Performance and Safety Benchmarking, in *2021 IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems (IROS)* **2021**, pp. 4290–4297.

[28] K. Vlack, T. Mizota, N. Kawakami, K. Kamiyama, H. Kajimoto, S. Tachi, Gelforce: A Vision-Based Traction Field Computer Interface, in *CHI'05 extended abstracts on Human factors in computing systems* **2005**, pp. 1154–1155.

[29] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, et al., Digit: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor with Application to In-Hand Manipulation, *IEEE Robot. Autom. Lett.* **2020**, *5*, 3838.

[30] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, S. Levine, Omnitact: A Multi-Directional High-Resolution Touch Sensor, in *2020 IEEE Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2020**, pp. 618–624.

[31] W. K. Do, B. Jurewicz, M. Kennedy, Densetact 2.0: Optical Tactile Sensor for Shape and Force Reconstruction, in *2023 IEEE Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2023**, pp. 12549–12555.

[32] J. Xu, W. Chen, H. Qian, D. Wu, R. Chen, Thintact: Thin Vision-Based Tactile Sensor by Lensless Imaging, *IEEE Trans. Robot.* **2025**, *41*, 1139.

[33] I. H. Taylor, S. Dong, A. Rodriguez, Gelslim 3.0: High-Resolution Measurement of Shape, Force and Slip in a Compact Tactile-Sensing Finger, in *2022 Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2022**, pp. 10781–10787.

[34] D. F. Gomes, Z. Lin, S. Luo, Geltip: A Finger-Shaped Optical Tactile Sensor for Robotic Manipulation, in *2020 IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems (IROS)* **2020**, pp. 9903–9909.

[35] A. Agarwal, M. A. Mirzaee, X. Sun, W. Yuan, A Modularized Design Approach for Gelsight Family of Vision-Based Tactile Sensors, *Int. J. Robot. Res.* **2025**.

[36] A. Agarwal, A. Wilson, T. Man, E. Adelson, I. Gkioulekas, W. Yuan, Vision-Based Tactile Sensor Design Using Physically Based Rendering, *Commun. Eng.* **2025**, *4*, 21.

[37] W. Friedl, M. A. Roa, Experimental Evaluation of Tactile Sensors for Compliant Robotic Hands, *Front. Robot. AI* **2021**, *8*, 704416.

[38] C. Higuera, A. Sharma, C. K. Bodduluri, T. Fan, P. Lancaster, M. Kalakrishnan, M. Kaess, B. Boots, M. Lambeta, T. Wu, M. Mukadam, Sparsh: Self-supervised touch representations for vision-based tactile sensing, **2024**. [Online]. Available: https://openreview.net/forum?id=xYJn2e1uu8.

[39] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, A. Rodriguez, 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), in *2018 IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems (IROS)*, IEEE **2018**, pp. 1927–1934.

[40] J. Lee, C. Son, J. Kim, D. Kang, S. Park, S. Jung, S. Ahn, C. Lim, D. Lee, D. Kim, J. Kim, H. Ko, C. Ryu, D. Baek, G. Jung, Y. Jeong, J.-H. Park, J. Eom, Behavioral Biometric Optical Tactile Sensor that Instantaneously Decouples Dynamic Touch Signals in Real Time, *Nat. Commun.* **2024**, *15*, 9079.

[41] W. Fan, H. Li, D. Zhang, Magictac: A Novel High-Resolution 3D Multi-Layer Grid-Based Tactile Sensor, in *2024 IEEE Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2024**, pp. 388–394.

[42] H. Li, Y. Lin, C. Lu, M. Yang, E. Psomopoulou, N. Lepora, Classification of Vision-Based Tactile Sensors: A Review, *IEEE Sens. J.* **2025**, *PP*, 1.

[43] Z. Kappassov, J.-A. Corrales, V. Perdereau, Tactile Sensing in Dexterous Robot Hands, *Robot. Auton. Syst.* **2015**, *74*, 195.

[44] W. K. Do, M. Kennedy, Densetact: Optical Tactile Sensor for Dense Shape Reconstruction, in *2022 Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2022**, pp. 6188–6194.

[45] C. Lin, H. Zhang, J. Xu, L. Wu, H. Xu, 9dtact: A Compact Vision-Based Tactile Sensor for Accurate 3D Shape Reconstruction And Generalizable 6d Force Estimation, *IEEE Robot. Autom. Lett.* **2023**.

[46] Z. He, X. Zhang, S. Jones, S. Hauert, D. Zhang, N. F. Lepora, Tacmms: Tactile mobile manipulators for warehouse automation, *IEEE Robot. Autom. Lett.* **2023**, *8*, 4729.

[47] G. Li, S. Liu, L. Wang, R. Zhu, Skin-Inspired Quadruple Tactile Sensors Integrated on a Robot Hand Enable Object Recognition, *Sci. Robot.* **2020**, *5*, eabc8134.

[48] W. Mandil, V. Rajendran, K. Nazari, A. Ghalamzan-Esfahani, Tactile-Sensing Technologies: Trends, Challenges and Outlook in Agri-Food Manipulation, *Sensors* **2023**, *23*, 7362.

[49] C. Lin, Z. Lin, S. Wang, H. Xu, Dtact: A Vision-Based Tactile Sensor that Measures High-Resolution 3D Geometry Directly From Darkness, in *2023 IEEE Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2023**, pp. 10359–10366.

[50] I. Andrussow, H. Sun, K. J. Kuchenbecker, G. Martius, Minsight: A Fingertip-Sized Vision-Based Tactile Sensor for Robotic Manipulation, *Adv. Intell. Syst.* **2023**, *5*, 2300042.

[51] H. Sun, K. J. Kuchenbecker, G. Martius, A Soft Thumb-Sized Vision-Based Sensor with Accurate All-Round Force Perception, *Nat. Mach. Intell.* **2022**, *4*, 135.

[52] X. Lin, L. Willemet, A. Bailleul, M. Wiertlewski, Curvature Sensing with a Spherical Tactile Sensor Using the Color-Interference of a Marker Array, in *2020 IEEE Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2020**, pp. 603–609.

[53] L. Zhang, Y. Wang, Y. Jiang, Tac3d: A Novel Vision-Based Tactile Sensor for Measuring Forces Distribution and Estimating Friction Coefficient Distribution, *arXiv preprint arXiv:2202.06211* **2022**.

[54] G. Zhang, Y. Du, H. Yu, M. Y. Wang, Deltact: A Vision-Based Tactile Sensor Using a Dense Color Pattern, *IEEE Robot. Autom. Lett.* **2022**, *7*, 10778.

[55] W. Kim, W. D. Kim, J.-J. Kim, C.-H. Kim, J. Kim, Uvtac: Switchable uv marker-based tactile sensing finger for effective force estimation and object localization, *IEEE Robot. Autom. Lett.* **2022**, *7*, 6036.

[56] S. Athar, G. Patel, Z. Xu, Q. Qiu, Y. She, Vistac Towards a Unified Multi-Modal Sensing Finger for Robotic Manipulation, *IEEE Sens. J.* **2023**.

[57] F. R. Hogan, M. Jenkin, S. Rezaei-Shoshtari, Y. Girdhar, D. Meger, G. Dudek, Seeing Through Your Skin: Recognizing Objects with a Novel Visuotactile Sensor, in *Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision* **2021**, pp. 1218–1227.

[58] S. Zhang, Y. Sun, J. Shan, Z. Chen, F. Sun, Y. Yang, B. Fang, Tirgel: A Visuo-Tactile Sensor with Total Internal Reflection Mechanism for External Observation and Contact Detection, *IEEE Robot. Autom. Lett.* **2023**.

[59] S. Quan, X. Liang, H. Zhu, M. Hirano, Y. Yamakawa, Hivtac: A High-Speed Vision-Based Tactile Sensor for Precise and Real-Time Force Reconstruction with Fewer Markers, *Sensors* **2022**, *22*.

[60] Q. Wang, Y. Du, M. Y. Wang, Spectac: A Visual-Tactile Dual-Modality Sensor Using UV Illumination, in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, **2022**, pp. 10844–10850.

[61] M. Pu, T. Zhao, L. Zhang, C. Han, Z. Chai, Y. Zhou, H. Ding, Z. Wu, An AI-Enabled All-In-One Visual, Proximity, and Tactile Perception Multimodal Sensor, *Adv. Robot. Res.* **2025**, *n/a*, 202500062.

[62] W. Fan, H. Li, D. Zhang, Magicgripper: A Multimodal Sensor-Integrated Gripper for Contact-Rich Robotic Manipulation, *arXiv preprint arXiv:2505.24382* **2025**.

[63] S. Zhang, Y. Yang, Y. Sun, N. Liu, F. Sun, B. Fang, Artificial Skin Based on Visuo-Tactile Sensing For 3D Shape Reconstruction: Material, Method, and Evaluation, *Adv. Funct. Mater.* **2025**, *35*, 2411686.

[64] K. Sato, K. Kamiyama, N. Kawakami, S. Tachi, Finger-Shaped Gelforce: Sensor for Measuring Surface Traction Fields For Robotic Hand, *IEEE Trans. Haptics* **2009**, *3*, 37.

[65] D. Ma, E. Donlon, S. Dong, A. Rodriguez, Dense Tactile Force Estimation Using Gelslim and Inverse Fem, in *2019 Inter. Conf. on Robotics and Automation (ICRA)*, IEEE **2019**, pp. 5418–5424.

[66] Y. Yang, X. Wang, Z. Zhou, J. Zeng, H. Liu, An Enhanced Fingervision for Contact Spatial Surface Sensing, *IEEE Sens. J.* **2021**, *21*, 16492.

[67] A. Yamaguchi, C. G. Atkeson, Tactile Behaviors with the Vision-Based Tactile Sensor Fingervision, *Int. J. Human. Robot.* **2019**, *16*, 1940002.

[68] Y. Du, G. Zhang, Y. Zhang, M. Y. Wang, High-Resolution 3-Dimensional Contact Deformation Tracking for Fingervision Sensor with Dense Random Color Pattern, *IEEE Robot. Autom. Lett.* **2021**, *6*, 2147.