MagicGripper: A Mini-MagicTac Integrated Gripper Enabling Multimodal Perception in Contact-Rich Manipulation

Wen Fan, Haoran Li, Qingzheng Cong, and Dandan Zhang

***Abstract*—Contact-rich robotic manipulation in unstructured environments demands reliable multimodal perception. Here, we present MagicGripper, a multimodal robotic gripper built around mini-MagicTac, a compact variant of the MagicTac sensor. Mini-MagicTac embeds multi-layer grid structures in a 3D-printed elastomer, enabling visual, proximity, and tactile sensing in a gripper-compatible form factor. In this paper, we introduce the design and multimodal perception capabilities of mini-MagicTac, as well as two algorithmic frameworks for proximity and contact detection. Experimental evaluations show that mini-MagicTac achieves high spatial resolution, accurate contact localisation, and robust force estimation under mechanical and manufacturing variations. Autonomous grasping trials further validate MagicGripper's reliable multimodal perception and adaptability to complex manipulation scenarios. These results demonstrate MagicGripper as a compact and versatile platform for embodied intelligence in contact-rich environments.**

***Note to Practitioners*—Robotic end-effectors often break down when a task calls for both "eyes" and "skin": adding multiple sensors usually makes the gripper bulky, fragile, and expensive to build. MagicGripper shows one practical way around that trade-off. Each finger is 3D-printed without casting or post-assembly is required; inside the soft skin a multi-layer grid acts as sensing feature, letting embedded camera read visual, proximity, and tactile cues simultaneously.**

***Index Terms*—Vision-based tactile sensor, multi-modality sensing, robotic manipulation.**

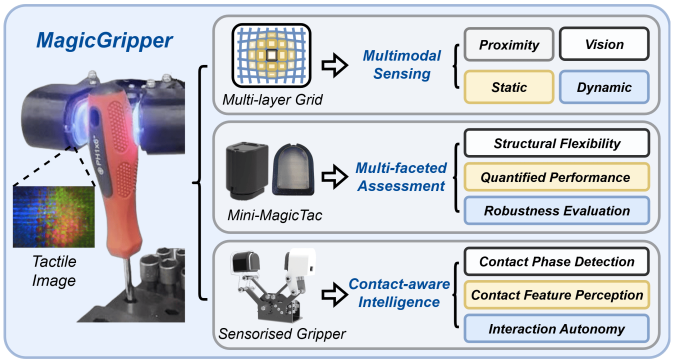


Fig. 1. Overview of the MagicGripper system featuring multimodal sensing via multi-layer grid structures. The integrated mini-MagicTac sensor provides visual, proximity, and tactile feedback, validated through extensive experiments in contact-aware manipulation.

I. Introduction

ROBOTIC manipulation in unstructured environments demands advanced sensing to perceive and respond to complex physical interactions between the robot and its surroundings [1], [2]. Contact-rich tasks particularly rely on *multimodal feedback* [3], [4], [5], combining visual, tactile, force, and proximity cues. Integrating these modalities into robotic end-effectors is essential for achieving human-like dexterity and adaptive interaction [6], [7], [8]. Among them, *vision* effectively captures object size, colour, shape, and texture [9], but its reliability decreases upon contact due to occlusion and lighting variation. In contrast, *tactile sensing* provides direct physical feedback and remains robust under visually degraded conditions. Tactile features can be categorised as *static* (e.g., texture, contact shape, indentation) and *dynamic* (e.g., force, shear, torque), jointly offering grounded information crucial for manipulation and exploration. Beyond vision and touch, *proximity sensing* bridges the perception-action gap by establishing a buffer zone for early surface detection and trajectory adjustment, enhancing precision and safety during fine manipulation. Hence, compact, compliant end-effectors capable of fusing these modalities are vital for robust contact-rich manipulation.

Among existing tactile sensing technologies, *vision-based tactile sensors (VBTSs)* have shown great promise for capturing rich contact information through embedded visual systems. However, current VBTS designs still face key trade-offs upon multimodal sensing that limit their use in compact end-effectors. Most VBTSs rely on specific optical encoding schemes that either offer high-resolution imaging but less dynamic force sensing, or vice versa. Moreover, modality switching between visual, proximity, and tactile cues often depends on additional hardware or complex calibration procedures, which increase system size and reduce robustness. These challenges, coupled with multi-stage fabrication and

Received 12 May 2025; revised 18 September 2025; accepted 20 October 2025. Date of publication 11 November 2025; date of current version 27 November 2025. The work of Wen Fan was supported by the China Scholarship Council (CSC). This article was recommended for publication by Associate Editor X. Li and Editor H. Moon upon evaluation of the reviewers' comments. *(Wen Fan and Haoran Li contributed equally to this work.) (Corresponding author: Dandan Zhang.)*

Wen Fan and Dandan Zhang are with the Department of Bioengineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: d.zhang17@imperial.ac.uk).

Haoran Li is with the School of Robotics, Xi'an Jiaotong-Liverpool University, Suzhou 215000, China.

Qingzheng Cong is with the Centre for Advanced Robotics (ARQ), Queen Mary University of London, E1 4NS London, U.K.

This article has supplementary downloadable material available at https://doi.org/10.1109/TASE.2025.3631485, provided by the authors.

Digital Object Identifier 10.1109/TASE.2025.3631485

limited structural flexibility, hinder seamless integration of VBTSs into multimodal robotic grippers.

To address these challenges, we introduce **MagicGripper** (Fig. 1), a compact, sensor-integrated gripper built on **mini-MagicTac**, a miniatured variant fabricated via multi-material additive manufacturing [10]. It employs a multi-layer grid elastomer architecture enabling simultaneous proximity, visual, and tactile perception without extra hardware for modality switching. The main contributions are as follows:

- **Sensor Design:** Development of mini-MagicTac, featuring an embedded multi-layer grid that enhances static and dynamic tactile sensing with design flexibility, manufacturing efficiency, and product consistency.
- **Multimodal Integration:** Realisation of unified visual, proximity, and tactile sensing by exploiting the reflective and refractive properties of grid cells, achieving multimodal perception within a single compact unit.
- **Sensing Framework:** Formulation of a perception framework that decouples proximity and contact events via channel entropy correlation and grid similarity metrics, enabling smooth transitions across pre-contact, contact, and post-contact phases.

Rather than proposing a single optimised design, MagicGripper demonstrates how multimodal perception enhances robotic interaction autonomy while offering a scalable pathway toward sensor-rich manipulators capable of contact-aware intelligence through flexible fabrication.

II. Related Work

A. Sub-Modal Tactile Sensing in VBTSs

Sub-modal tactile sensing refers to a sensor's capability to capture distinct tactile components such as static texture or dynamic force. In vision-based tactile sensors (VBTSs), this capability depends on the sensing mechanism [11], which determines how physical interactions are optically encoded into images. A widely adopted approach is the *intensity mapping method (IMM)* used in GelSight-type designs [12], [13], [14], [15], where reflective surface coatings record static features including fine geometry and local indentation. In contrast, the *marker displacement method (MDM)* tracks the motion of embedded markers to sense dynamic contact information [16]. Most *MDM*-based VBTSs employ a single marker layer, either embedded within the elastomer [17] or patterned on its surface [18]. Multi-layer configurations extend this principle: GelForce [19] uses dual marker layers to estimate traction fields, while ChromaTouch [20], [21] applies subtractive colour mixing for richer tactile cues. Hybrid *IMM-MDM* designs [9], [22], [23], [24] enable simultaneous perception of static and dynamic features but often inherit the drawbacks of both methods. Reflective coatings are susceptible to abrasion, and dense marker patterns obscure optical cues, reducing spatial fidelity. UVtac [25] mitigates this trade-off through controllable ultraviolet (UV) illumination switching.

B. Multi-Modality Sensing of VBTSs

Beyond improving sub-modal sensing, the *modality fusion method (MFM)* [11] integrates complementary modalities such as vision, proximity, and thermal cues within a unified framework to enhance perception. *MDM-MFM* systems tend to utilize transparent elastomers to combine visual and tactile sensing, like FingerVision [26], [27] and ViTacTip [28]. ViTacTip further employs a generative adversarial network (GAN) for data-level modality switching. SpecTac [29] incorporates switchable UV markers and transparent layers to capture both contact and spectral visual information. *IMM-MFM* systems achieve modality switching via optical path control. STS [30] and VisTac [31] employ internal lighting and mirror coatings to alternate between visual and tactile modes, whereas TIRgel [32] uses focus adjustment for mode transition. Beyond visual-tactile fusion, SATac [33] integrates a thermoluminescent layer for simultaneous tactile-thermal sensing, and M3Tac [34] combines near- and mid-infrared imaging for multispectral perception of texture, force, proximity, and temperature. Recently, DIGIT360 [35] introduced an omnidirectional, high-resolution fingertip capable of sensing normal and shear forces, vibration, temperature, and chemical feedback. These existing MFM-based systems demonstrate the potential of cross-modal fusion to broaden the perception bandwidth. However, their designs remain complex, bulky, and hardware-intensive, limiting scalability and hindering seamless integration into robotic grippers for contact-rich manipulation.

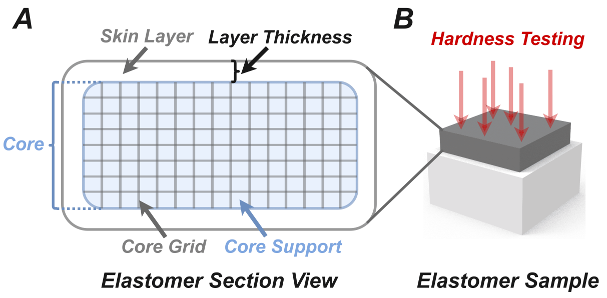


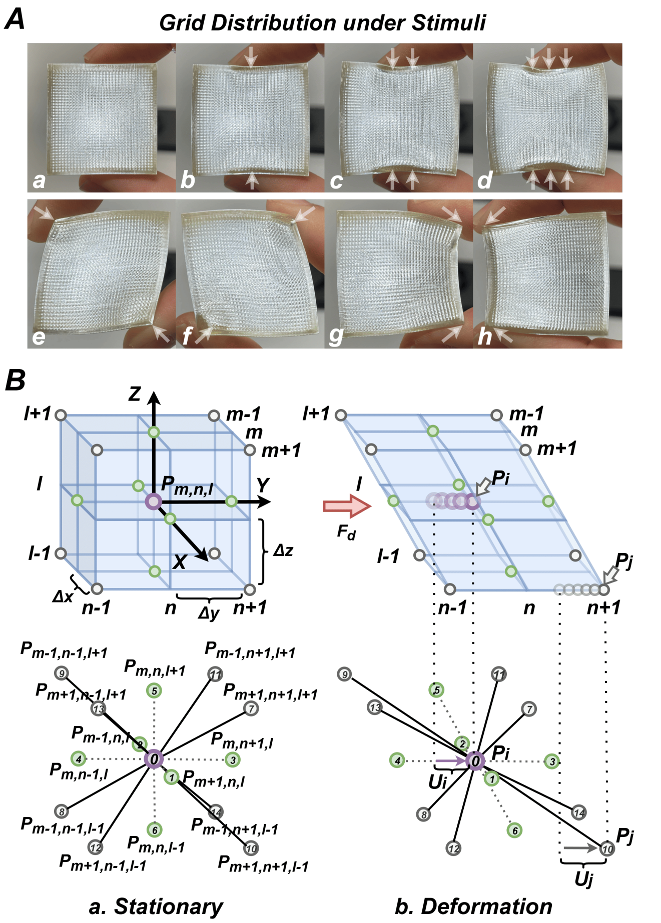
Fig. 2. Printed elastomer in mini-MagicTac with embedded multi-layer grid. A: External skin and internal core structure. B: Samples for hardness test.

III. Methodology

This section outlines design of mini-MagicTac, the core of MagicGripper. The underlying sensing mechanism is first detailed, followed by its integration into MagicGripper.

A. Multi-Layer Grid Embedded in Mini-MagicTac

As described in [10], the multi-layer grid is fabricated through PolyJet multi-material additive manufacturing. This integrated process simultaneously prints internal structures with different materials, eliminating post-print assembly and overcoming the design and quality limitations of manual fabrication. As shown in Fig. 2(A), the elastomer comprises an embedded core and an external skin. Grid-cell dimensions, typically 0.6-1 mm, can be adjusted to balance deformability and printing precision. The grid skeleton, printed using Agilus30 Clear for its rubber-like elasticity and durability, is filled with support material SUP706 to enhance integrity and reduce hardness. Both Agilus30 Clear (transparent) and SUP706 (translucent) transmit light, forming the basis for



a. Stationary b. Deformation

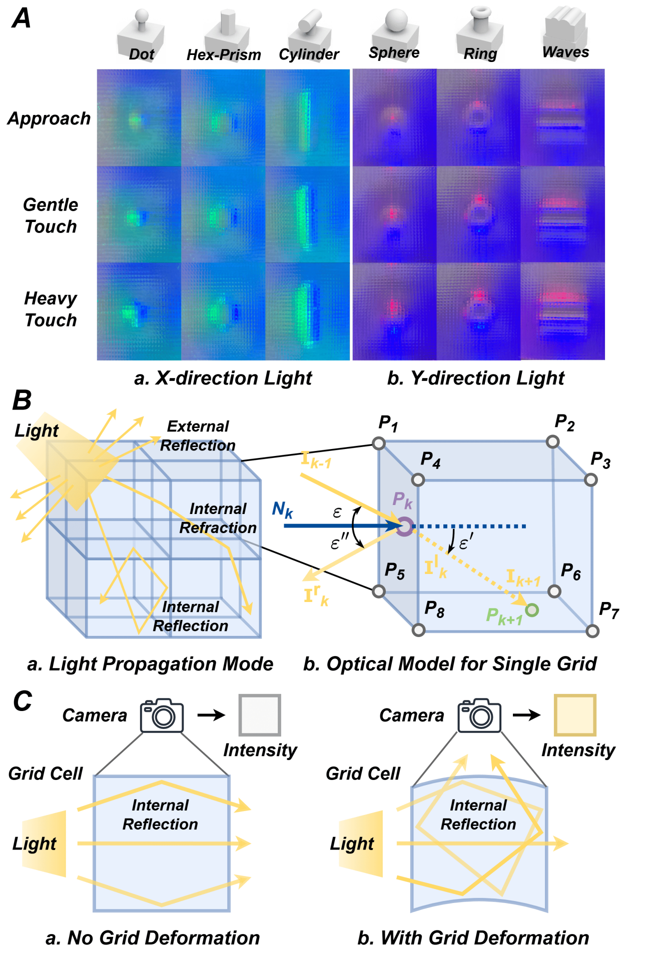
Fig. 3. A: Observed grid distributions near stimuli showing sensitivity to force magnitude (b-d), direction (e-f), and rotation (g-h). B: Lattice spring model of the multi-layer grid in (a) stationary and (b) deformed states.

multimodal sensing. The co-printed external skin uniformly encloses the structure. Material choice, Agilus30 Clear or Agilus30 Black, determines optical behaviour: the transparent variant maximises light transmission, whereas the opaque variant blocks external illumination. Skin thickness, adjustable from 0.3 mm, further tunes hardness through geometry. The performance of this design is evaluated in Section IV-A using the test samples shown in Fig. 2(B).

B. Sensing Properties of the Multi-Layer Grid

1) Deformation Analysis: As shown in Fig. 3(A), the spatial distribution of grid cells changes systematically with stimulus magnitude and direction. Without contact, the grid remains uniform (Fig. 3(A.a)); under light load, only the surface layer deforms elastically along the applied force (Fig. 3(A.b)). As the load increases, (*i*) deformation magnitude grows proportionally and (*ii*) the affected region expands outward from the contact zone (Fig. 3(A.c-d)). These relationships persist under changes in force direction and rotation (Fig. 3(A.e-h)), indicating isotropic response.

To interpret this behaviour, a lattice spring model (LSM) is introduced (Fig. 3(B)). The representative structure consists of a 2 × 2 × 2 array of flexible cells $(\Delta x, \Delta y, \Delta z)$ distributed across three layers: upper $(l+1)$, middle (l), and lower $(l-1)$. These eight cells define the *representative elementary volume* (REV) of the grid, centred at node $P_{m,n,l}$ (purple). In the stationary state (Fig. 3(B.a)), four structural nodes (green) and eight diagonal nodes (white) are connected by linear springs. When a contact force F_d is applied, connected nodes (P_i, P_j) experience displacements (u_i, u_j) (Fig. 3(B.b)), whose magnitude and orientation depend on F_d. Multiple REVs distributed throughout the elastomer form a continuous mechanical field, enabling fine-grained sub-modal tactile capture.



a. X-direction Light b. Y-direction Light

a. Light Propagation Mode b. Optical Model for Single Grid

a. No Grid Deformation b. With Grid Deformation

Fig. 4. A: Optical responses of multi-layer grid under various case. B: Three light propagation modes coupled within the grid and (b) optical model for each cell describing light transport. C: Deformation concentrates light within cells through internal reflection, increasing brightness in the camera image.

2) Optical Analysis: The deformed grid exhibits distinct optical behaviour under varying contact conditions. As shown in Fig. 4(A), faint contours of an object appear as it approaches due to partial light scattering. Upon contact, cells in direct contact brighten sharply; as indentation deepens, the grid conforms to surface curvature, expanding the illuminated region. These stage-wise intensity changes are consistent across different objects and lighting conditions.

An optical model is formulated based on the REV in the LSM (Fig. 4(B)). A light ray incident at a random angle may undergo three propagation modes: (*i*) external reflection, (*ii*) internal reflection, and (*iii*) internal refraction.

- **External and Internal Reflection:** When the incident angle ϵ of incoming light I_{k-1} is large at cell sidewalls (P_1, P_4, P_5, P_8), only part I_{k+1} enters the cell,

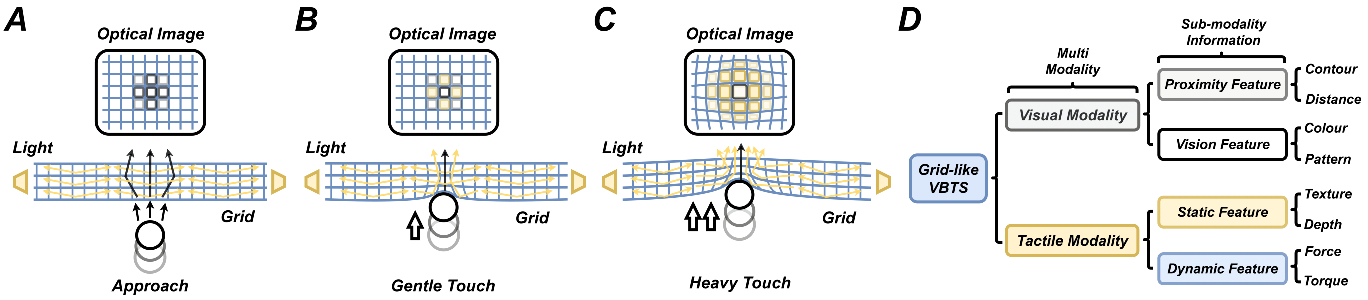


Fig. 5. Multimodal sensing principle of the multi-layer grid. A: During object approach, the grid remains undeformed while proximity and visual cues are transmitted through internal refraction. B: Under light contact, only the surface deforms locally, producing internal reflection that brightens contact regions (static tactile features). C: With heavier contact, deformation propagates through multiple layers, capturing dynamic tactile features such as force and torque. D: Grid-like VBTS fuses visual and tactile modalities, encompassing sub-modal information of proximity, visual detail, and static/dynamic contact cues.

while the remainder I_k^r is reflected outward at angle ϵ''. This becomes more pronounced under oblique illumination. Internal reflection allows light to propagate within deformed cells, concentrating brightness (Fig. 4(C)).

- **Internal Refraction:** At smaller angles, I_{k-1} passes through the sidewall with refracted angle ϵ', generating transmitted light I_k^l. Near-normal incidence increases transmission, enabling light to traverse neighbouring cells and reach the camera, contributing to intensity.

Overall, the grid-stimulus interaction can be modelled as soft-body deformation driven by contact forces. The LSM serves as a discrete analogue of finite element analysis (FEM), where grid-cell distribution represents the internal stress-strain field of the elastomer. Although direct tracking of 3D grid-node motion from 2D images remains challenging, the grid acts as a transparent elastomer embedded with translucent mesh segments. This structure permits partial light transmission and introduces discrete refraction and reflection within deformed regions, encoding geometric and intensity variations that underpin multimodal sensing in mini-MagicTac. The further theoretical model of above process has been proposed in Supplementary Section VII-A.

3) Multimodal Sensing Principle: Building on the deformation and optical analyses, the multimodal sensing principle of the multi-layer grid is summarised in Fig. 5. When illuminated by side-mounted LEDs and viewed by a top-mounted camera, an approaching object without contact (Fig. 5(A)) maintains a stable illumination pattern dominated by total internal reflection (TIR). Simultaneously, proximity and visual cues are captured through vertical internal refraction. At larger distances, **coarse proximity features** such as contours and approximate depth dominate; as the object nears, **fine visual details** such as texture and colour distribution emerge, defining the grid's **visual modality**.

Under light contact (Fig. 5(B)), the outer skin compresses locally, altering internal reflection and creating bright regions corresponding to **static tactile features** such as texture and indentation depth. With heavier force (Fig. 5(C)), deformation penetrates deeper layers, translating force and torque into internal stress and strain. These deformations appear as broad optical variations encoding **dynamic tactile features** such as force magnitude, direction, and motion.

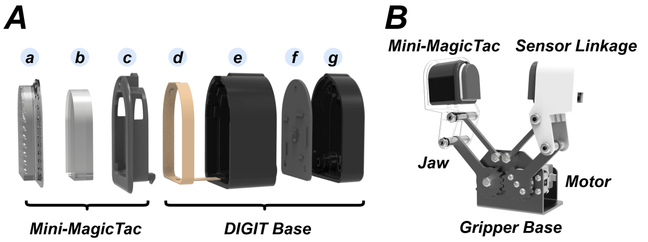


Fig. 6. Hardware design of the mini-MagicTac and MagicGripper. A: Exploded view of the mini-MagicTac assembly: (a) printed elastomer with embedded multi-layer grid, (b) sensor lens, (c) sensor base, (d) LED strip, (e) DIGIT upper base, (f) camera, and (g) DIGIT lower base. B: MagicGripper adopts a two-finger, motor-driven design where each finger integrates a mini-MagicTac module for multimodal perception.

The integrated framework (Fig. 5(D)) fuses visual and tactile modalities. The visual modality dominates pre-contact and near-contact phases, providing proximity and appearance cues, whereas the tactile modality activates upon contact, capturing static and dynamic interaction features. Grid density determines tactile resolution: smaller cells enhance texture sensitivity, while excessive layering reduces optical clarity. Conversely, too few layers restrict deformation along the Z-axis, improving visual sharpness but reducing force and depth perception. Hence, grid thickness (the number of layers for a given density) must be optimised to balance optical transmission and elasticity, as discussed in [11].

C. Hardware Design of MagicGripper

To demonstrate the multimodal sensing capability in real-world manipulation, we propose the hardware design of MagicGripper, illustrated in Fig. 6. Leveraging the design flexibility of integral printing [10], the multi-layer grid architecture can be adapted to various VBTS configurations and robotic systems. The mini-MagicTac, a compact finger-shaped sensing module compatible with the DIGIT base unit,[1] is shown in Fig. 6(A). The module consists of a printed elastomer with an embedded multi-layer grid, an optical lens, a camera, and side-mounted LED illumination within DIGIT housing. Its small form factor ensures compatibility with robotic end-effectors while maintaining high-resolution sensing.

[1] https://github.com/facebookresearch/digit-design

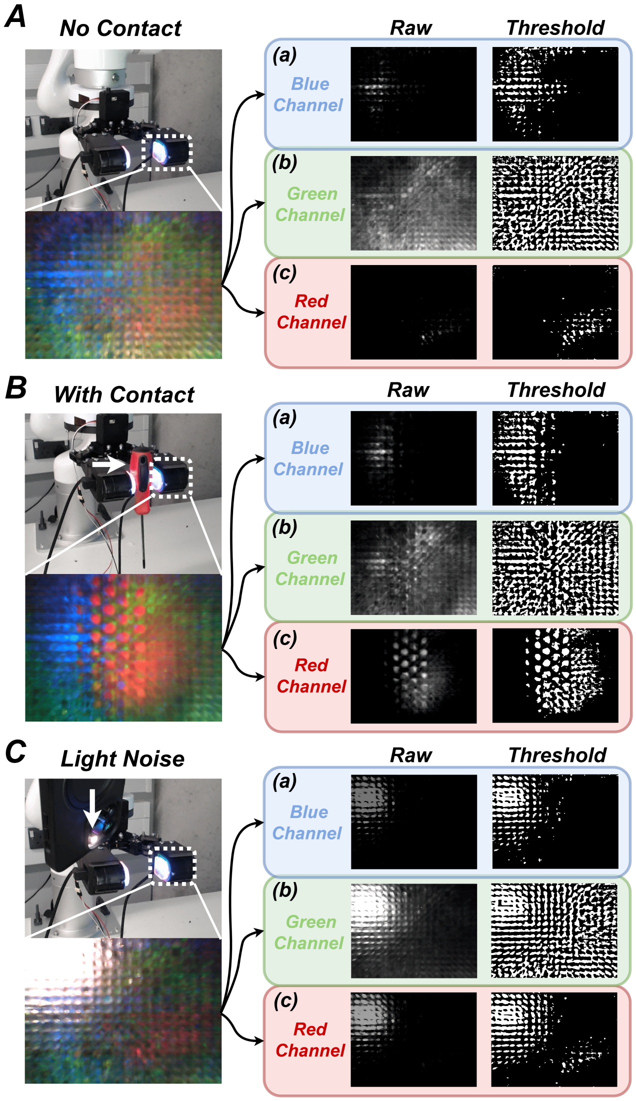


Fig. 7. Sensing property analysis of mini-MagicTac. A/B: Under heterochromatic illumination from different angles, each RGB channel captures distinct features depending on proximity or contact state. C: External lighting interference reduces inter-channel contrast, degrading signal reliability.

Building on this design, the MagicGripper adopts a two-finger architecture in which each motor-driven finger integrates one mini-MagicTac module (Fig. 6(B)). A base-mounted motor drives both fingers symmetrically to open and close, maintaining parallel alignment of the sensing surfaces for stable, contact-rich grasps. During operation, each finger captures multimodal sensory data, which are fused to achieve closed-loop manipulation.

D. Proximity and Contact Detection Algorithm

Building on the multimodal sensing principle of the multi-layer grid (Fig. 5), proximity and contact detection algorithms were developed for MagicGripper. These algorithms enable responsive, contact-aware manipulation by exploiting the channel-wise optical characteristics of mini-MagicTac.

1) Channel Response Characterization: To analyse the sensing behavior of the mini-MagicTac mounted on the DIGIT base, the RGB channel responses were analysed under heterochromatic illumination (Fig. 7), which reveal distinct channel characteristics. The blue and red channels exhibit uneven illumination, whereas the green channel provides uniform coverage across the field of view (Fig. 7(A)).

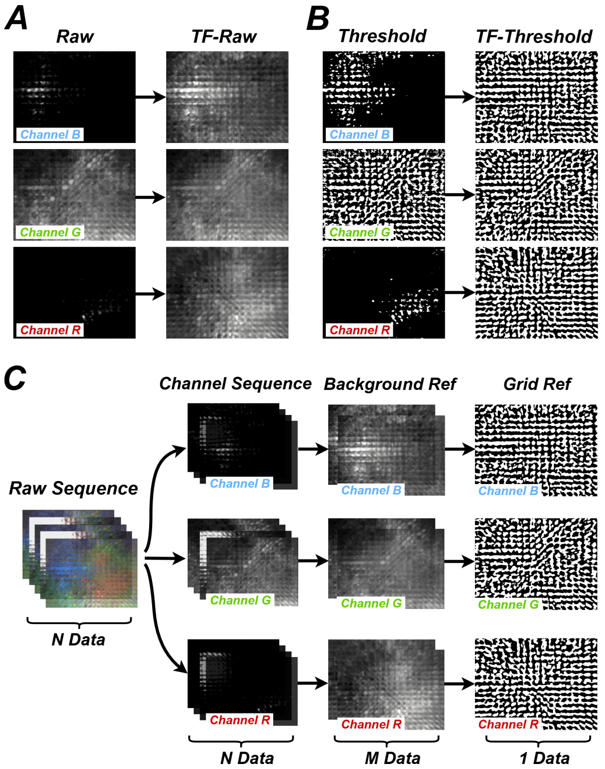


Fig. 8. A/B: Temporal fusion (TF) enhances both background suppression and global grid recognition. C: For reference generation, TF is applied channel-wise to raw images using an $N:M$ ratio, producing M background references and one grid reference.

Channel response varies with contact state and external lighting. Under strong ambient interference (Fig. 7(C)), differences between channels diminish, obscuring proximity cues and reducing data reliability. In contrast, during contact (Fig. 7(B)), inter-channel contrast increases, highlighting deformation and grid illumination. Threshold-based analysis further shows that the grid geometry appears most distinct in the green channel, demonstrating high optical sensitivity. However, cross-channel variance and sensitivity to light noise remain challenges for stable signal extraction.

2) Temporal Fusion Method for Data Processing: To extract stable and informative features from each colour channel, a **temporal fusion (TF)** method is introduced. It averages raw image data over a defined period to generate representative reference frames, typically during the initialisation phase of mini-MagicTac. As illustrated in Fig. 5(A), although illumination may vary due to internal reflection and refraction, the geometry of the embedded grid remains static in the absence of contact. This implies that image data contain consistent patterns over time, especially in grid-associated regions.

Experimental results support this observation (Fig. 8(A,B)). The visual outputs (**TF-raw** and **TF-threshold**) exhibit stable global features and well-defined grid structures, confirming the effectiveness of the temporal fusion strategy. The TF-raw image, produced by averaging raw frames, preserves the overall illumination pattern while suppressing high-frequency noise

Algorithm 1 Temporal Fusion for Data Processing

```
Input:
$I = \{I_1, \ldots, I_N\}$ : sequence of raw images
$N = 30$ : number of raw images
$M = 3$ : number of background reference masks
$\tau_B = 35$ : binary threshold for grid reference mask
Output:
$B_{ref} = \{B_1, \ldots, B_M\}$: background reference masks
$G_{ref}$ : grid reference mask

Step 1: Channel Decomposition
for $I_i \in I$ do
    Split $I_i$ into channels $c \in \{r, g, b\}$: $I_i^r$, $I_i^g$, $I_i^b$

Step 2: Background Mask Construction
for $j = 1$ to $M$ do
    for channel $c \in \{r, g, b\}$ do
        $B_j^c = \text{mean}(\{I^c_{(j-1)\cdot(N/M)+1}, \ldots, I^c_{j\cdot(N/M)}\})$
    $B_j = \text{merge}(\{B_j^r, B_j^g, B_j^b\})$
$B_{ref} = \{B_1, \ldots, B_{j=M}\}$

Step 3: Grid Mask Construction
for channel $c \in \{r, g, b\}$ do
    for $i = 1$ to $N$ do
        Apply Gaussian blur to $I_i^c$ to filter noise
        $G_i^c = \text{binary threshold}(I_i^c, \tau_B)$
    $G^c = \text{mean}(\{G_1^c, \ldots, G_{i=N}^c\})$
$G_{ref} = \text{merge}(\{G^r, G^g, G^b\})$
return $B_{ref}$, $G_{ref}$
```

Algorithm 2 Proximity Detection Using Channel Entropy and Inter-Channel Correlation

```
Input:
$F_{\text{curr}}$ : current image frame
$\tau_E = 0.5$ : channel entropy threshold
$\tau_C = 0.2$ : channel correlation threshold
Output:
Proximity state $\in$ {Normal, Approaching, Noise}

Step 1: Background Mask Generation
Same as Step 2 in Algorithm 1:
$B_{ref} = \{B_1, \ldots, B_M\}$: background reference masks
$M = 3$ : number of background reference masks

Step 2: Preprocess Current Frame
Split $F_{\text{curr}}$ into channels $c \in \{r, g, b\}$: $F^r$, $F^g$, $F^b$
for channel $c \in \{r, g, b\}$ do
    for $j = 1$ to $M$ do
        $\Delta F_j^c = \text{threshold}(F^c - B_j^c, 0)$
        $F_j^c = F_{j-1}^c \cap \Delta F_j^c$
    $F^c_{filtered} = F^c_M$
$F_{intersection} = \text{merge}(\{F^r_{filtered}, F^g_{filtered}, F^b_{filtered}\})$

Step 3: Compute Entropy and Correlation
for channel $\{r, g, b\}$ do
    $C_{rg} = \text{correlation}(F^r_{filtered}, F^g_{filtered})$
    $C_{rb} = \text{correlation}(F^r_{filtered}, F^b_{filtered})$
    $C_{gb} = \text{correlation}(F^g_{filtered}, F^b_{filtered})$
$E_{\text{total}} = \text{entropy}(gray(F_{intersection}))$
$C_{\text{total}} = \text{mean}(C_{rg}, C_{rb}, C_{gb})$

Step 4: State Decision Logic
if $E_{total} < \tau_E$ then
    return Normal
else if $E_{total} \geq \tau_E \boldsymbol{and} C_{total} < \tau_C$ then
    return Approaching
else if $E_{total} \geq \tau_E \boldsymbol{and} C_{total} \geq \tau_C$ then
    return Noise
```

from ambient light fluctuations or minor vibrations. In contrast, TF-threshold applies binary thresholding before fusion, enhancing high-contrast regions, sharpening grid edges, and improving structural segmentation.

Based on these results, two reference masks are constructed from the temporal sequence of mini-MagicTac images: a **background reference mask** and a **grid reference mask**. The background reference mask filters environmental noise, whereas the grid reference mask provides a baseline for detecting deformation during contact. As shown in Fig. 8(C) and Algorithm 1, a set of N raw images is first split by RGB channels. Each channel is processed with a temporal fusion ratio of $N : M$ to generate M background reference masks (where $M \leq N$). The parameters N and M are user-defined and can be tuned according to lighting variability, enhancing robustness to temporal disturbances. The grid reference mask, by contrast, is generated as a single fused image from the entire N-frame sequence, ensuring maximal retention of the static grid geometry and serving as a reliable reference for downstream tasks such as contact detection.

3) Algorithm Framework Design: Algorithmic frameworks for proximity and contact detection are developed to address two key challenges: (1) distinguishing valid proximity cues from mini-MagicTac's raw data beside light noise, and (2) extracting contact features without grid tracking.

a) Proximity detection: As shown in Fig. 9(A), the raw image sequence first undergoes background denoising and cumulative channel intersection (Fig. 9(A.a-b)). This identifies consistent variations across channels over M iterations, corresponding to the number of reference masks. The filtered data (F^r, F^g, F^b) are then processed by two modules: entropy computation and inter-channel correlation analysis. Channel entropy E quantifies visual disorder as:

$$E = -\sum_i P(i) \log_2 P(i) \tag{1}$$

where $P(i)$ is the normalised histogram of pixel intensities. Higher E values indicate increased image complexity, often caused by the approach of an external object. Since entropy is sensitive to illumination noise, it is complemented by a correlation-based stability measure. Pearson's correlation coefficient between channels F^x and F^y is given by:

$$C_{xy} = \frac{cov(F^x, F^y)}{\sigma_{F^x}\sigma_{F^y}} \tag{2}$$

where $cov(F^x, F^y)$ is the covariance and $\sigma_{F^x}, \sigma_{F^y}$ are standard deviations. Iterating through all channel pairs yields a correlation matrix describing inter-channel relations.

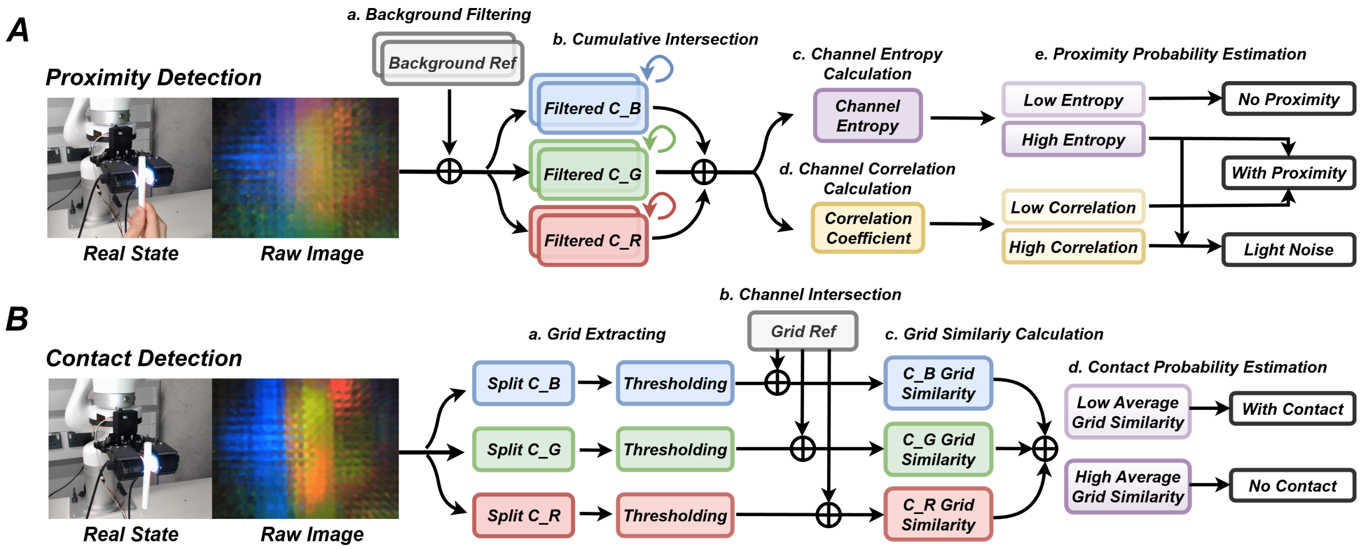


Fig. 9. Proximity and contact detection frameworks. A: **Proximity detection:** (a) remove background using reference frames; (b) perform cumulative channel intersection, followed by (c) entropy and (d) correlation computation to estimate proximity probability. B: **Contact detection:** (a) extract grid structure from each channel; (b) apply grid references for intersection; (c) compute grid similarity for contact probability estimation.

An increase in entropy accompanied by a decrease in correlation indicates an approaching object, as internal reflection and refraction alter heterochromatic illumination differently across channels. Conversely, high correlation during entropy variation suggests global lighting noise. Stable low entropy corresponds to a non-contact state (Fig. 9(A.e)). The complete implementation is summarised in Algorithm 2.

b) Contact detection: The contact detection pipeline, shown in Fig. 9(B), focuses on grid deformation analysis. As observed in Fig. 5(B-C), heavier contact induces greater deformation within the multi-layer elastomer. Each RGB channel of the mini-MagicTac output is thresholded to extract the intrinsic grid geometry (Fig. 9(B.a)), which is compared with the reference grid pattern generated through temporal fusion (Algorithm 1, Fig. 8). The ratio of overlapping pixels between the current grid and reference defines the *grid similarity*:

$$S = \frac{A_{\text{intersect}}}{A_{\text{reference}}} \tag{3}$$

A higher S indicates minor deformation, whereas lower S corresponds to stronger contact (Fig. 9(B.c-d)). The detection logic is summarised in Algorithm 3.

Algorithm 3 Contact Detection Using Grid Similarity

```
Input:
$F_{\text{curr}}$ : current image frame
$G_{\text{ref}}$ : pre-computed grid reference mask
$\tau_B = 35$ : binary threshold for grid reference mask
$\tau_G = 0.6$ : grid similarity threshold
Output:
Contact state $\in$ {Touched, Untouched}

Step 1: Preprocess Current Frame
Split $F_{\text{curr}}$ into channels $c \in \{r, g, b\}$: $F^r$, $F^g$, $F^b$
for channel $\{r, g, b\}$ do
    Apply Gaussian blur to $F^c$ to filter noise
    $G^c = \text{binary threshold}(F^c, \tau_B)$

Step 2: Grid Mask Intersection
Split $G_{\text{ref}}$ into channels $c \in \{r, g, b\}$: $G^r_{ref}$, $G^g_{ref}$, $G^b_{ref}$
for channel $c \in \{r, g, b\}$ do
    $G^c_{intersection} = G^c_{ref} \cap G^c$

Step 3: Compute Grid Similarity
for channel $\{r, g, b\}$ do
    $S_r = similarity(G^r_{intersection}, G^r_{ref})$
    $S_g = similarity(G^g_{intersection}, G^g_{ref})$
    $S_b = similarity(G^b_{intersection}, G^b_{ref})$
$S_{\text{total}} = \text{mean}(S_r, S_g, S_b)$

Step 4: State Decision Logic
if $S_{total} < \tau_G$ then
    return Touched
else if $S_{total} \geq \tau_G$ then
    return Untouched
```

IV. Experimental Evaluation

This section presents the experimental evaluation of both mini-MagicTac and MagicGripper. The former is assessed in terms of design flexibility, sensing performance, and robustness, while the latter is evaluated for contact-phase detection, contact-feature perception, and interaction autonomy.

A. Flexibility Evaluation of Multi-Layer Grid

To quantify the tunable mechanical properties of the multi-layer grid, two experiments were conducted using the printed samples in Fig. 2(B). Hardness was measured with a Shore-A durometer at five evenly distributed surface points.

In the first test, elastomer height was fixed at 10 mm while the external skin thickness varied from 0.5 mm to 3

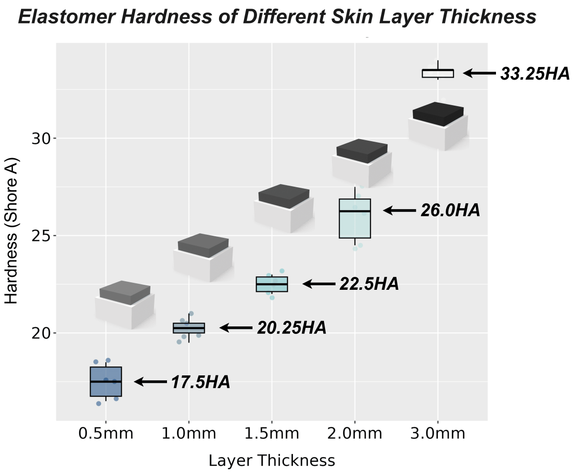


Fig. 10. Hardness of multi-layer grid can be modified between 17A and 33A when skin layer of elastomer samples varies from 0.5mm to 3mm thick.

TABLE I

PRINTED ELASTOMERS WITH ADJUSTABLE STRUCTURES & HARDNESS

Elastomer Height	Skin Thickness	Core Height	Test Hardness
10mm	3mm	4mm	33.25A
10mm	2mm	6mm	26.00A
10mm	1.5mm	7mm	22.25A
10mm	1mm	8mm	20.25A
10mm	0.5mm	9mm	17.50A
2mm	0.5mm	1mm	30.50A
5mm	0.5mm	4mm	18.20A
15mm	0.5mm	14mm	15.17A
20mm	0.5mm	19mm	14.20A

mm, reducing the grid-core height from 9 mm to 4 mm. In the second, skin thickness remained 0.5 mm and total height increased from 2 mm to 20 mm, yielding a proportional core height from 1 mm to 19 mm.

From Fig. 10, hardness increased linearly from 17.5A to 33.25A as skin thickness rose from 0.5 mm to 3 mm, approaching the value of pure Agilus30 Clear (35A). Table I indicates that, under fixed skin thickness, greater core height produces softer samples. With a 0.5 mm skin, increasing the core height from 1 mm to 19 mm reduced hardness from 30.5A to 14.2A. For reference, typical VBTS elastomers measure 5A-20A [12]. By adjusting skin thickness and core height, the grid achieves a broader range (15A-30A) than pure Agilus30, offering greater design tunability for VBTS.

Design constraints remain: very thin skins, though softer, are prone to abrasion under repeated contact, while thicker skins improve durability but reduce grid-core height and tactile sensitivity. Balancing compliance and robustness, a skin thickness of 0.5-1 mm and a core height of about 5 mm are recommended for the embedded grid in mini-MagicTac.

B. Performance Evaluation of Mini-MagicTac

1) Spatial Resolution: Spatial resolution is a key indicator of static tactile performance in VBTSs. As shown in Fig. 5(A,B), the multi-layer grid captures surface texture through visual cues and internal reflection within its cells. To validate this capability, a comparative study was conducted between mini-MagicTac and the commercial GelSight sensor, a recognised benchmark for high-resolution tactile imaging. Two calibration samples, a *dot array* (convex) and a *line grille* (concave), were fabricated, each containing 25 configurations with feature sizes from 0.20 mm to 1.75 mm in 0.05 mm increments (Fig. 11). These primitives represent common texture geometries, allowing comprehensive evaluation.

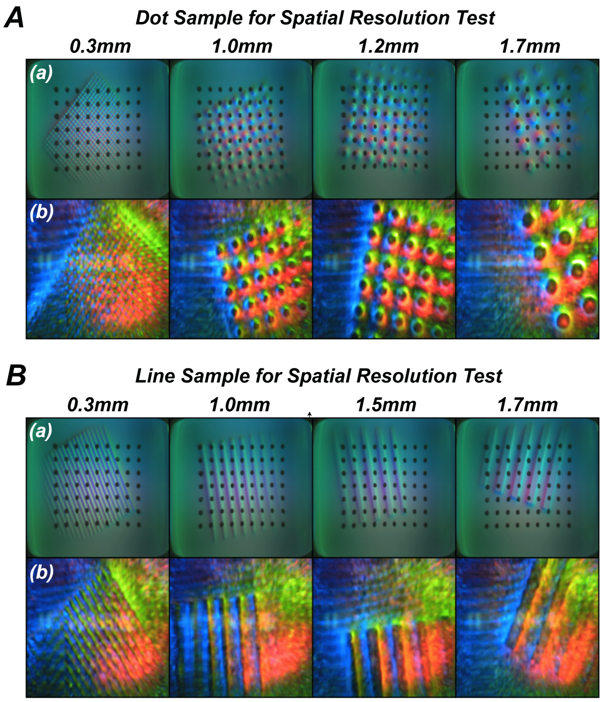

A
Dot Sample for Spatial Resolution Test
0.3mm
1.0mm
1.2mm
1.7mm
(a)
(b)
B
Line Sample for Spatial Resolution Test
0.3mm
1.0mm
1.5mm
1.7mm
(a)
(b)

Fig. 11. A/B: Dot and line samples used for spatial resolution testing of GelSight (a) and mini-MagicTac (b), with various feature dimensions.

TABLE II

SENSING ACCURACY (%) AT DIFFERENT SPATIAL RESOLUTION (*mm*)

Sensor	0.50	0.30	0.25	0.20	0.15	0.10	0.05
GelSight (dot)	100	100	100	100	100	100	99.03
GelSight (line)	100	100	100	100	100	100	99.72
MagicTac (dot)	100	100	100	100	100	99.67	98.33
MagicTac (line)	100	100	100	100	100	99.00	98.00

Both sensors were mounted on a robotic arm performing controlled indentations with randomised position (x, y), depth (0.1-2 mm), and rotation angle ($-90°$ to $90°$). Each configuration was repeated 100 times, producing 2500 samples per sensor under identical conditions. Separate ResNet-18 models were trained for each dataset using cross-entropy loss, with RGB tactile images as input and the 25 texture sizes as categorical outputs. Data were split into training, validation, and test subsets (7:2:1), and training ran for 100 epochs with early stopping and learning-rate decay.

Spatial resolution was defined as the minimum feature spacing where two textures could be distinctly recognised. Classification accuracy was evaluated within tolerance windows of $\pm\Delta$ mm (Δ = 0.05-0.50 mm). Predictions within

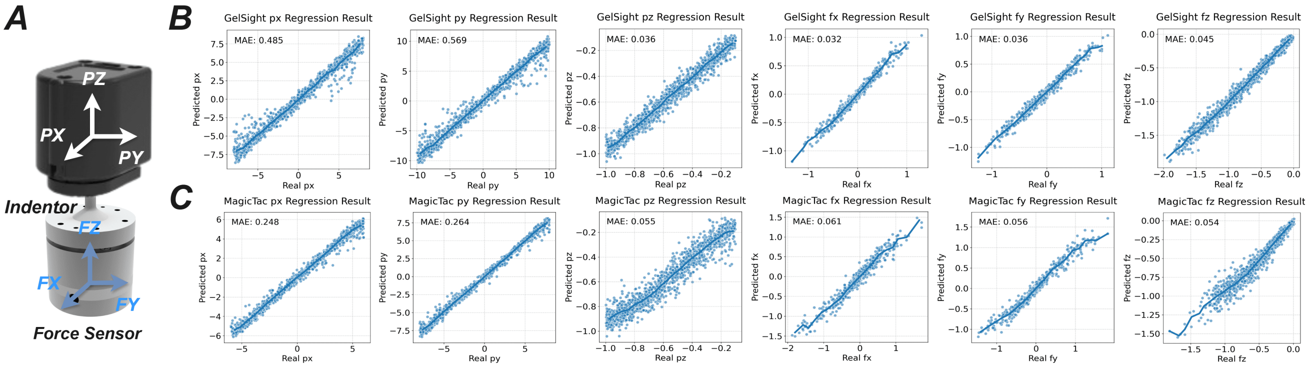


Fig. 12. A: Experimental setup for contact localization and force regression, where six pose and force are recorded. B/C: Results of GelSight/mini-MagicTac.

tolerance were counted as correct, and confusion matrices were generated to compute recognition accuracy.

Results in Table II show that both sensors achieved 100% accuracy for separations above 0.15 mm. When feature spacing fell below this value, mini-MagicTac accuracy declined, while GelSight maintained stable recognition down to 0.10 mm before degrading. Hence, mini-MagicTac attains a spatial resolution of approximately 0.15 mm, comparable to GelSight.

2) Contact Localization and Force Regression: Accurate estimation of contact position and force is critical for dynamic tactile sensing. As illustrated in Fig. 5(C), deformation of the multi-layer grid allows mini-MagicTac to capture both spatial and force-related cues. To evaluate its dynamic performance, a comparative experiment was conducted using the setup in Fig. 12(A), where the sensor mounted on a robotic arm contacts an indenter placed above a force sensor. The ground-truth force was recorded by the force sensor, and contact position was derived from the arm's pose.

Each trial comprised a sequence of controlled indentations. The sensor pressed from the initial contact to a random depth between 0.1 mm and 2 mm, discretized into four levels. At each depth, normal and shear data were acquired by laterally translating the sensor up to 1 mm in the XY-plane while maintaining contact. The process was repeated four times, ensuring that identical spatial locations were revisited under varying normal- or shear-dominant forces, enriching the dataset. The training configuration followed that of the spatial resolution task but reformulated as a regression problem using mean absolute error (MAE) loss. RGB tactile images served as input, and the network predicted both contact location (p_x, p_y, p_z) and force (f_x, f_y, f_z), with results in Fig. 12(B,C).

For localisation, mini-MagicTac achieved sub-millimetre accuracy in the XY-plane around 0.25mm, surpassing GelSight's 0.5mm. Along the Z-axis, its error was slightly higher than GelSight's, likely due to GelSight's reflective surface providing stronger depth contrast. For force estimation, GelSight yielded smaller errors around 0.04N than mini-MagicTac at 0.05N. Overall, mini-MagicTac demonstrates competitive dynamic sensing, achieving sub-millimetre contact localisation and sub-0.1 N force estimation within a compact form factor.

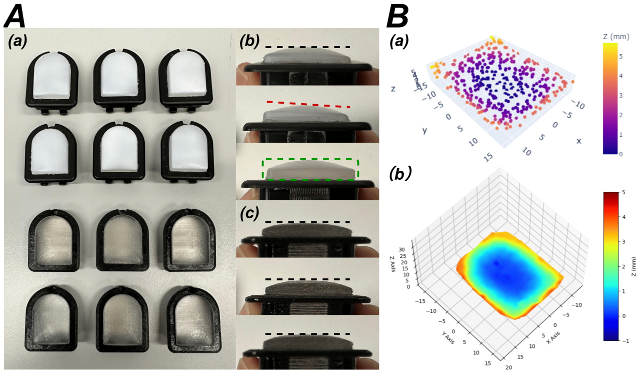


Fig. 13. A: (a) Twelves DIGITs and mini-MagicTacs prepared for manufacturing error evaluation; (b) Significant variation observed in the manufacturing of DIGITs; (c) Mini-MagicTac exhibits more consistent manufacturing quality. B: (a) Multi-point random sampling conducted on each sensor's surface to obtain a point cloud; (b) Final reconstruction results of the sensor surface.

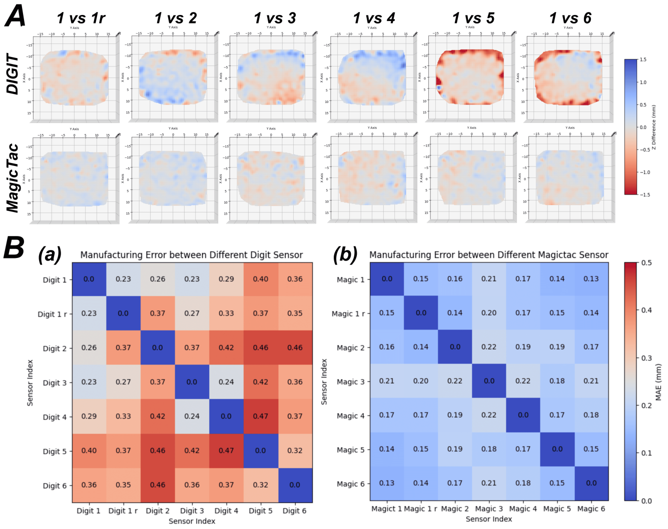


Fig. 14. A: Surface error distribution maps for DIGIT (top) and mini-MagicTac (bottom), where "1r" indicates self-repeat tests. B: Manufacturing error matrices of (a) DIGIT and (b) mini-MagicTac.

C. Robustness Evaluation of Mini-MagicTac

To demonstrate MagicGripper's practicality, its embedded mini-MagicTac was firstly assessed in manufacturing, mechanical, performance, and interference robustness.

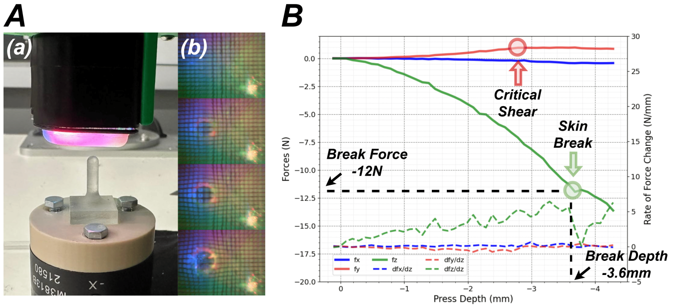


Fig. 15. A: Mechanical robustness evaluation through destructive skin puncture testing: (a) test setup and (b) sequence of indenter piercing the skin. B: Force-depth curves showing the elastic-plastic transition and failure stages.

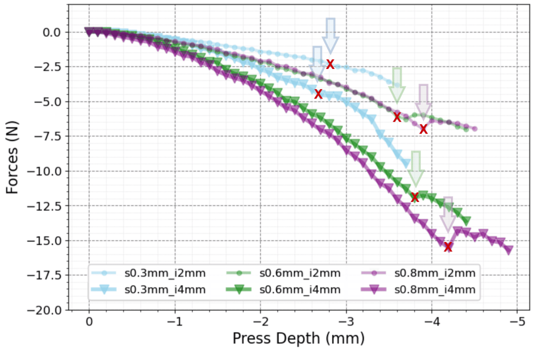


Fig. 16. Skin puncture test results under varying skin thickness and indenter diameters. The red cross marks the rupture point; 'sXmm_iYmm' denotes skin thickness X mm and indenter diameter Y mm.

TABLE III
MECHANICAL ROBUSTNESS TEST OF MINI-MAGICTAC WITH DIFFERENT SKIN THICKNESS AND INDENTOR SIZE

Sensor	Critical Shear	Break Force	Break Depth
s0.3mm_i2mm	-2.0N	-2.25N	-2.8mm
s0.3mm_i4mm	-3.5N	-4.5N	-2.7mm
s0.6mm_i2mm	-3.7N	-6.25N	-3.6mm
s0.6mm_i4mm	-7.0N	-12.0N	-3.8mm
s0.8mm_i2mm	-3.7N	-7.0N	-3.9mm
s0.8mm_i4mm	-7.2N	-15.5N	-4.4mm

Note: 'sXmm_iYmm' refers to skin thickness (X) and indentor diameter (Y)

1) Manufacturing Robustness: Using integral multi-material printing [10], each mini-MagicTac is fabricated as a single unit. To evaluate fabrication quality, a comparative study was conducted with DIGIT [14], another widely used VBTS with prevalence in tactile research, employing conventional assembly but sharing the same mechanical base as mini-MagicTac.

As shown in Fig. 13(A.a), six DIGIT units from three production batches were examined to minimise batch bias, and six mini-MagicTacs were printed using DIGIT's CAD model for direct comparison. DIGIT sensors presented two recurring issues: (*i*) variation in elastomer surface slope (red line, Fig. 13(A.b)), causing inconsistent contact geometry, and (*ii*) non-uniform elastomer thickness (green box), leading to mismatch between indentation depth and tactile signal. In contrast, mini-MagicTac exhibited consistent surface profiles across units (black curves, Fig. 13(A.c)). To quantify deviations, sensor surfaces were reconstructed over a square region using the same setup as the force regression test. Valid contact points were detected via thresholding, with 400 samples collected per unit (Fig. 13(B.a)) and reconstructed through noise filtering and surface fitting (Fig. 13(B.b)).

Error maps in Fig. 14(A) reveal that DIGIT units 1/2 show minor deviations, 3/4 display skew in the upper-right area, and 5/6 exhibit thicker surfaces. Conversely, all mini-MagicTacs demonstrate uniform geometry. Global results (Fig. 14(B)) show similar self-repeat errors for both sensors (0.15-0.23 mm), representing baseline reconstruction uncertainty. However, DIGIT's inter-unit deviation ranges above to 0.4 mm, while mini-MagicTac maintains less than 0.2 mm.

These results confirm that integral printing yields higher structural consistency and geometric precision for mini-MagicTac than conventional VBTS fabrication, improving reproducibility and scalability.

2) Mechanical Robustness: The mechanical robustness of mini-MagicTac determines the long-term stability of MagicGripper during physical interaction. It depends mainly on two parameters: (1) outer-skin thickness and (2) pressure sustained by the internal support material within the multi-layer grid.

Structural robustness was evaluated via a puncture test (Fig. 15(A.a)) using a domed cylindrical indenter under the same setup as Fig. 12(A). During indentation, the indenter first deforms the skin and underlying grid. At a critical depth, the support material in the affected cells yields plastically, and further pressing ruptures the skin (Fig. 15(A.b)), where the white region marks support failure following piercing. In Fig. 15(B), when f_x or f_y reaches a shear threshold, a transition can be found from an increasing to a plateau trend, leading to shifts from elastic to plastic deformation. At this point, f_z indicates the maximum safe load before damage. Further increasing f_z causes irreversible fracture of support followed by skin tearing, consistent with failure observation.

To assess robustness comprehensively, two indenters (diameters 2 mm and 4 mm) and three skin thicknesses (0.3, 0.6, 0.8 mm) were tested, as summarised in Fig. 16. For a given thickness, smaller indenters produced higher local stress and were more prone to puncture, while thicker skins withstood greater loads and indentation depths. However, internal damage often occurred before visible tearing, consistent with the critical transition in Fig. 15(B). As reported in Table III, f_z at the onset of shear was consistently lower than the final rupture force. With increasing skin thickness, the ratio of f_z (critical shear) to breaking force decreased from about 80% (0.3 mm) to 60% (0.6 mm) and 50% (0.8 mm), indicating stronger outer layers while the internal support remained unchanged.

Based on these findings, the recommended operating limits for mini-MagicTac are a maximum pressing force of 6 N and indentation depth of 2.5 mm. For contact with sharp or rigid objects, these should be conservatively limited to 3 N and 2 mm to ensure long-term mechanical reliability.

3) Performance Robustness: Unlike visual sensors, tactile sensors undergo frequent contact, leading to gradual wear that can degrade performance. Although mini-MagicTac's integral

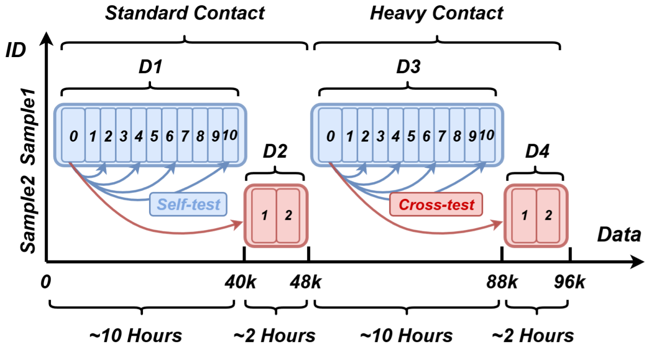


Fig. 17. Two mini-MagicTac samples continuously collected data for 24 hours to evaluate performance robustness. Four datasets (D1-D4) were obtained, totalling 96,000 data for self-testing and cross-testing.

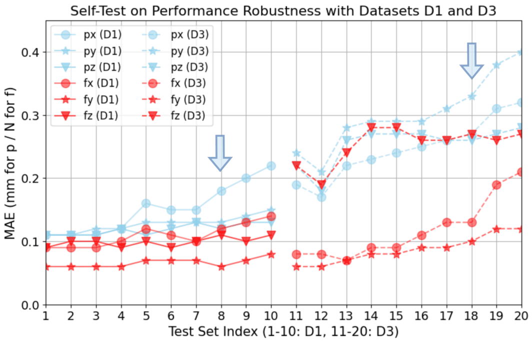


Fig. 18. Self-test results using sample 1. Under standard contact (D1), performance remains stable up to 32,000 data. A similar trend is observed under heavy contact (D3), with degradation occurring beyond 80% cycles.

fabrication minimises unit deviation, the generalisation of its tactile algorithms within MagicGripper requires validation. To this end, a long-duration wear test was conducted using the same setup as Fig. 12(A). A domed indenter identical to that in the puncture test (Fig. 15(A)) repeatedly rubbed the sensor surface under continuous contact. Based on Table III, a 6 mm indenter was selected for its larger area, reducing the likelihood of premature rupture and enabling extended testing.

To evaluate worst-case degradation, two mini-MagicTacs with the thinnest (0.3 mm) skin were tested. As shown in Fig. 17, four datasets were collected: D1 and D3 from sample 1, D2 and D4 from sample 2. Each dataset captured normal and shear interactions over 24 hours of continuous operation, totalling 96,000 samples. For standard contact (D1, D2), press depth and shear displacement were 1.2 mm and 1.0 mm; for heavy contact (D3, D4), they increased to 1.6 mm and 1.5 mm, corresponding to force range (f_x, f_y, f_z) of $(2, 2, 2)$N for standard contact and $(3, 3, 4)$N for heavy contact, respectively. D1 and D3 contained ~40,000 samples each, D2 and D4 ~8,000.

a) Self-test: The first 10% of D1 and D3 was used to pretrain the regression model (same configuration as Fig. 12); the remaining 90% was divided into ten temporally ordered test sets, with results shown in Fig. 18. Accuracy remained stable for the first 80% ($\leq$ 32,000 interactions) before a marked decline. D3, collected after D1 under heavier contact, exhibited

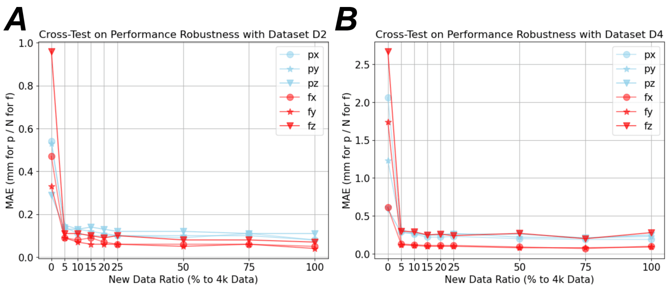


Fig. 19. Cross-test results using sample 2. Zero-shot transfer between mini-MagicTacs initially shows an accuracy gap, but fine-tuning with 5% new data improves generalisation under both (A) standard and (B) heavy contact.

roughly double the mean absolute error (MAE), particularly in f_z, due to cumulative wear and higher contact forces.

b) Cross-test: To assess cross-unit generalisation, models pretrained on D1/D3 (sample 1) were evaluated on D2/D4 (sample 2). Half of each dataset was reserved for testing, and the rest incrementally used for fine-tuning (5-100%), with 80/20 training-validation splits. As shown in Fig. 19, zero-shot transfer initially produced noticeable errors, reflecting minor inter-unit variation despite consistent fabrication. However, fine-tuning with only 5% of new data (~160 samples) rapidly restored performance, reducing MAE to ~0.1 for D2 and ~0.3 for D4, comparable to baseline results in Fig. 12.

Overall, mini-MagicTac maintained reliable sensing accuracy after extended wear and demonstrated efficient cross-unit adaptation with minimal recalibration, confirming both hardware and algorithmic robustness.

4) Light Interference Robustness: To rigorously evaluate the effect of ambient illumination on mini-MagicTac, two complementary experiments were conducted. First, **coloured light sources** were used to test the robustness of the proximity detection algorithm under variations in colour, intensity, and position-critical since the algorithm relies on decoupled RGB-channel responses. Second, **white light sources** with different intensities and directions were applied to examine their influence on force estimation, which depends mainly on data-driven learning and is therefore less sensitive to colour but more affected by illumination geometry.

a) Proximity light robustness: The setup (Fig. 20(A.a)) included a digital lux meter and a 16-colour LED strip, with available colours listed in Table IV. Two illumination modes were tested: (1) *Single-LED mode*, where one LED simulated a point-source interference (Fig. 20(A.b)); and (2) *Multi-LED mode*, where a coiled strip generated distributed lighting (Fig. 20(A.c)). Representative red (R0), green (G0), and blue (B0) examples are shown in Fig. 20(B). The measured illuminance of all 16 LEDs (Fig. 20(C)) showed that single LEDs produced weaker light than multi-LEDs, e.g., bright white (W) yielded ~1500 lx versus ~3000 lx. Red (R0-R4) ranged from 50-1300 lx, green (G0-G4) from 1300-2200 lx, and blue (B0-B4) remained stable around 1500-1700 lx, providing diverse colour and intensity conditions.

As shown in Fig. 9(A), proximity detection relies on channel entropy and inter-channel correlation to classify "no proximity," "proximity," and "light noise" states. Their

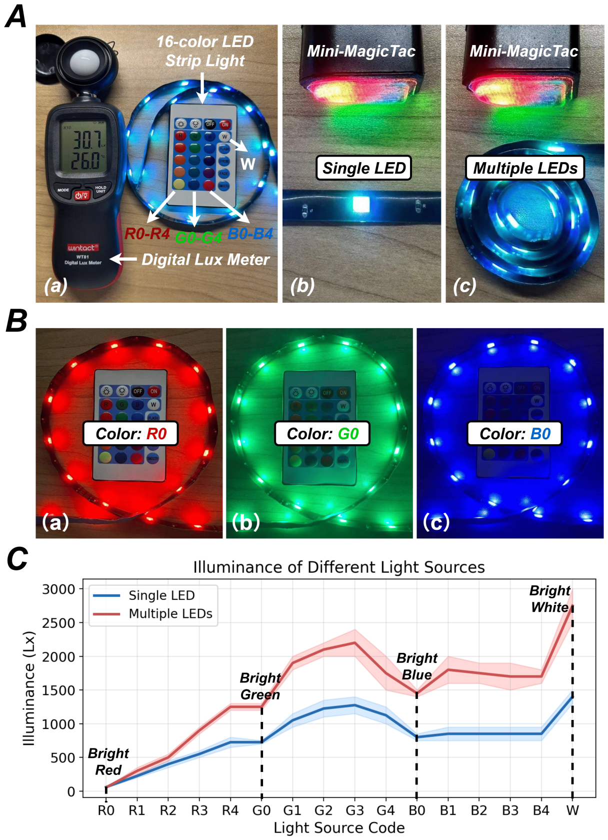


Fig. 20. A: Setup for light robustness evaluation: (a) digital lux meter and 16-colour LED strip, (b) single-LED mode, (c) multi-LED mode. B: Example LED colours: (a) bright red, (b) bright green, (c) bright blue. (C) Measured illuminance of 16 light sources at 1 mm intervals.

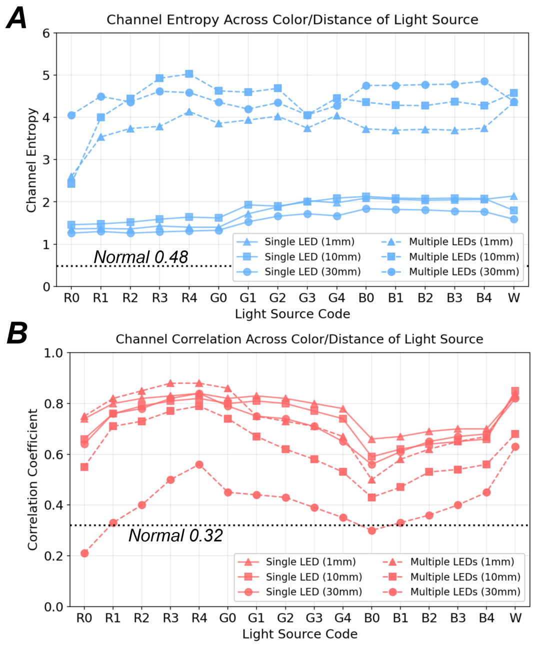


Fig. 21. A: Channel entropy of mini-MagicTac across 16 colours and three distances. B: Channel correlation across the same conditions.

TABLE IV
AVAILABLE COLOURS FOR LIGHT ROBUSTNESS EXPERIMENTS

No.	Colour Name	Code	No.	Colour Name	Code
1	Bright Red	R0	9	Aqua Cyan	G3
2	Orange Red	R1	10	Dark Cyan	G4
3	Bright Orange	R2	11	Bright Blue	B0
4	Orange Yellow	R3	12	Navy Blue	B1
5	Bright Yellow	R4	13	Violet Purple	B2
6	Bright Green	G0	14	Grape Purple	B3
7	Teal Green	G1	15	Magenta Purple	B4
8	Deep Teal	G2	16	Bright White	W

behaviour under varying illumination was analysed (Fig. 21). For channel entropy (Fig. 21(A)), multi-LED lighting consistently produced higher entropy than single LEDs, exceeding the indoor baseline of 0.48 due to increased geometric complexity and total intensity. The lowest entropy appeared at the smallest spacing (1 mm), confirming spatial complexity as the main driver. Among single LEDs, red light yielded slightly lower entropy than green or blue, and entropy decreased with source distance, indicating that intensity dominated effects.

For channel correlation (Fig. 21(B)), both illumination modes maintained values between 0.6-0.8, indicating general insensitivity to geometry and intensity. However, colour had a stronger influence: correlation decreased from red to green to blue, with bright blue (B0) showing the largest drop. Correlation also declined with increasing source distance, particularly in the multi-LED setup, where at 30 mm spacing, R0 and B0 correlations fell below the indoor baseline.

In summary, channel entropy is primarily governed by illumination geometry and intensity, while colour and distance play secondary roles. Conversely, channel correlation is more sensitive to colour and distance but largely invariant to geometry. Accordingly, mini-MagicTac's proximity robustness can be enhanced through adaptive threshold: reducing entropy threshold τ_E under single-point lighting and lowering correlation threshold τ_C under multi-point or blue-dominant illumination.

b) Tactile light robustness: To assess the effect of illumination on force estimation, four white-light configurations were tested (Fig. 22(A)): **LM1**, a diffuse ceiling light (80 lx); **LM2**, a wide-angle mobile flashlight (700 lx); **LM3**, a ring light below the sensor providing uniform circumferential illumination (1500 lx); and **LM4**, a narrow-angle directional source (1700 lx). Natural indoor light at 20 lx served as baseline. For comparison, ViTacTip [28], another VBTS with a transparent silicone skin, was evaluated in parallel.

Before regression, two metrics quantified lighting effects:(1) the **mean grayscale value**, reflecting overall image brightness, and(2) the **structural similarity index (SSIM)** [36], measuring spatial consistency between normal and perturbed lighting.Under natural light, ViTacTip exhibited a low grayscale mean of 11-13 (on [0,255]), while mini-MagicTac remained higher (100-115) due to its embedded RGB lighting and internal reflection.For each illumination mode, corresponding image pairs were matched via Euclidean distance in pose space (p_x, p_y, p_z) to compute SSIM.

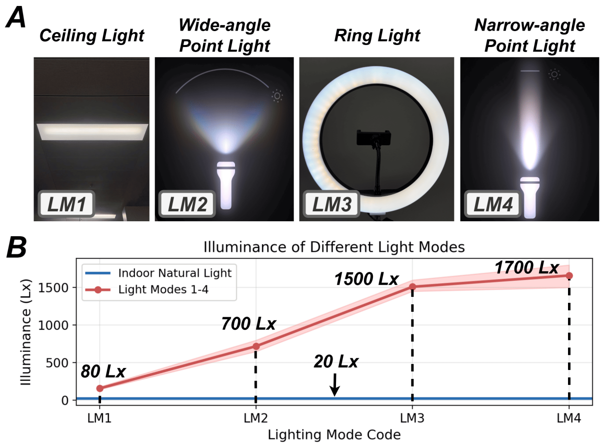


Fig. 22. A: Four lighting configurations: LM1 (ceiling light), LM2 (wide-angle point light), LM3 (ring light), LM4 (narrow-angle point light). B: Measured illuminance values compared with natural indoor lighting.

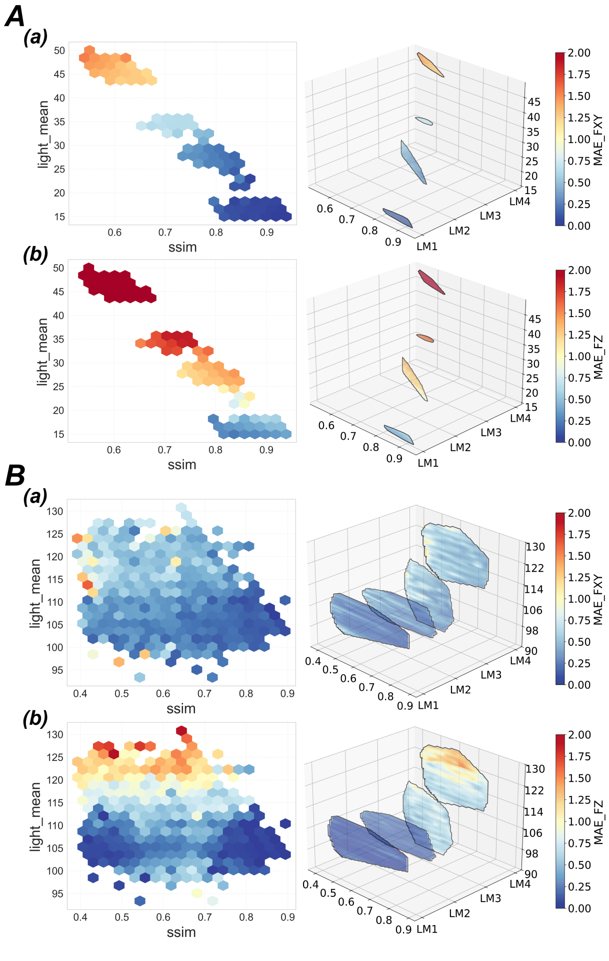


Fig. 23. A: Force regression performance of ViTacTip under different light modes. B: Performance of mini-MagicTac under the same conditions.

Force regression was evaluated in terms of mean grayscale, SSIM, and mean absolute error (MAE) for tangential (f_{xy}) and normal (f_z) forces (Fig. 23).For ViTacTip (Fig. 23(A)), both grayscale and SSIM varied linearly with light intensity: as brightness increased from LM1 to LM4, SSIM dropped steadily while grayscale rose from 15 to 50, reflecting its high transparency.Consequently, force accuracy declined markedly under strong illumination, especially along the Z-axis, where MAE for f_z reached 1 N under LM2 and up to 2 N under LM4; f_{xy} remained below 1 N for LM1-LM3.

TABLE V

Performance Evaluation Using Different Models

Test Error	Px	Py	Pz	Fx	Fy	Fz	Rz
Pretrained Model	1.12	0.16	0.06	0.07	0.06	0.12	1.84
New (no finetune)	2.26	1.40	0.43	0.82	0.49	1.85	12.8
New (finetuned)	1.18	0.15	0.08	0.18	0.12	0.28	2.46

In contrast, mini-MagicTac (Fig. 23(B)) showed minimal sensitivity to light variation.Its grayscale increased modestly from 100-110 (LM1-LM2) to 110-130 (LM4), without linear dependence.SSIM fluctuated between 0.4-0.9 with no monotonic trend, indicating weak sensitivity to illumination geometry.This stability arises from optical scattering and refraction within the grid cells (Fig. 4) and RGB backlighting from the DIGIT base (Fig. 7).Even at 1700 lx (LM4), force estimation remained stable: MAE for f_{xy} stayed below 0.75 N and f_z below 0.5 N for LM1-LM2, rising only to ~1.25 N at LM4-about half that of ViTacTip.

These results reveal two key factors: First, the multi-layer grid functions as an optical filter that attenuates external light fluctuations, unless illuminance exceeds a threshold. Second, isotropic reflection and refraction within the grid maintain image quality under varying light directions. Together, these properties enable mini-MagicTac to sustain stable in diverse lighting conditions, outperforming designs whose elastomer directly transmits environmental light.

D. Contact Alignment Task With MagicGripper

To validate the contact inference capability of mini-MagicTac shown in Fig. 12, a **contact alignment task** was conducted to assess MagicGripper's responsiveness.

1) Experiment Design: Six cylindrical parts (10 cm length; diameters 10, 15, and 20 mm, each in black and white) were 3D-printed for testing. During operation, MagicGripper sensed the rotational pose of the grasped cylinder and adjusted the robot arm orientation to maintain vertical alignment.

Three mini-MagicTacs were used: **Sensor1** and **Sensor2** were embedded in the left and right gripper fingers, while **Sensor3** collected large-scale data for model pretraining.It acquired 1,800 samples per object (10,600 total) labelled with contact position (p_x, p_y, p_z), rotation angle r_z, and forces (f_x, f_y, f_z).Data were split 7/2/1 for training, validation, and testing, following the same model as Fig. 12.

To examine model transferability, Sensors1/2 each collected 160 samples across three random objects. Half were used for testing, and the rest split 8/2 for fine-tuning and validation. As summarised in Table V, test loss decreased from 2.88 to 0.64 after fine-tuning, indicating that less than 5% new data suffices for substantial error reduction.

2) Experiment Analysis: The setup is shown in Fig. 24(A).Each mini-MagicTac outputs its local contact orientation r_z, but the two sensors are mounted oppositely, yielding mirrored readings. Taking Sensor1 as reference, Sensor2's

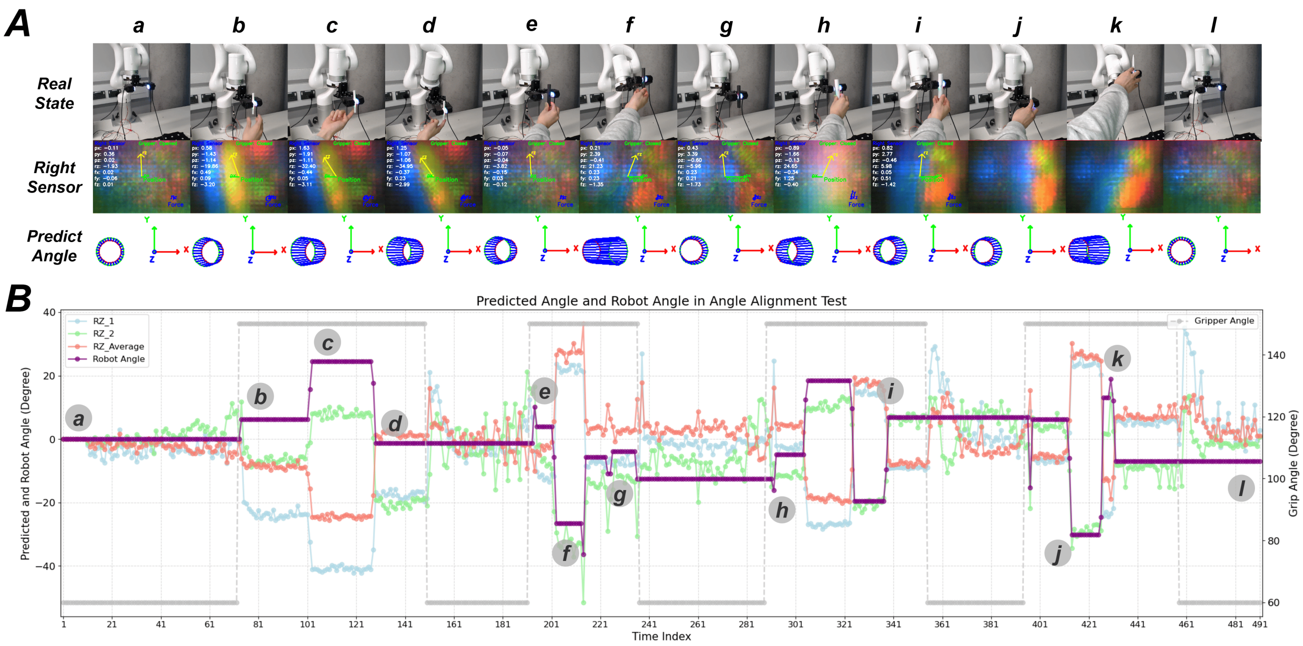


Fig. 24. A: Three 3D-printed cylindrical parts with varying diameters and colours were selected for testing, together with a pen as an additional object to evaluate generalisation. B: Real-time contact alignment framework of MagicGripper.

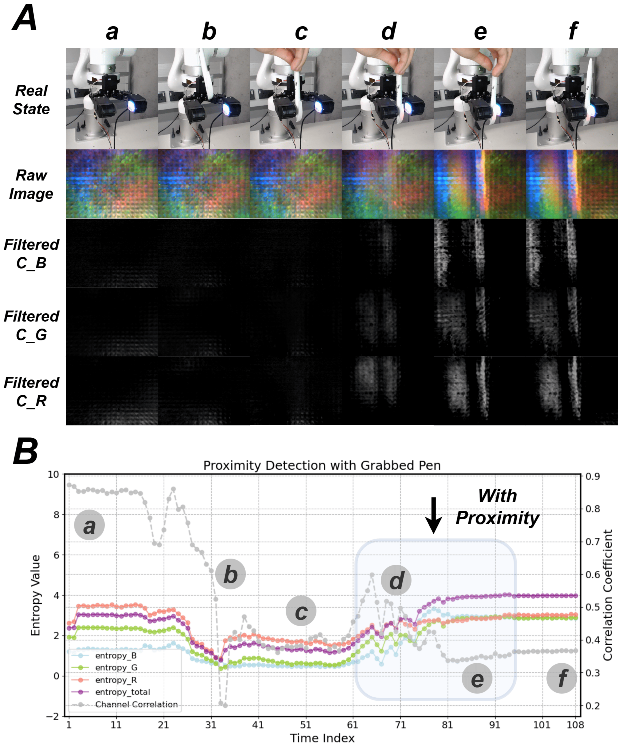


Fig. 25. A: Channel feature distributions deviate as an external object approaches (a-c). B: The proximity event leads to decreased inter-channel correlation and increased entropy due to changing visual cues (d-f).

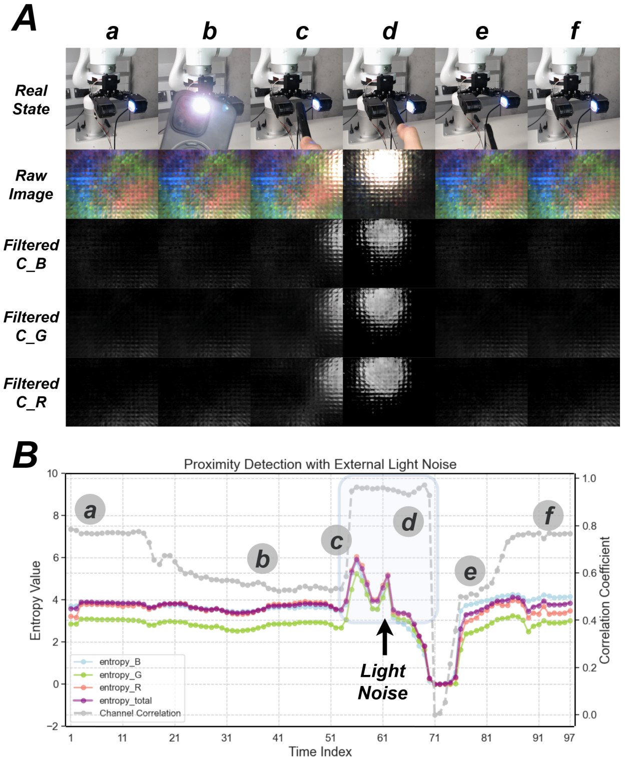


Fig. 26. A: External illumination uniformly affects all channels (c-d). B: During light interference, inter-channel correlations increase while channel entropies fluctuate upward due to illumination variation (c-d).

output is inverted as $-r_{z2}$, and the averaged contact orientation is computed as $r_{z,\text{avg}} = |r_{z1} - r_{z2}|/2$, while the robot's compensatory rotation is set to $r_{\text{robot}} = -r_{z,\text{avg}}$.It enables MagicGripper to maintain object alignment perpendicular.

Three representative parts—a 10 mm white, 15 mm black, and 20 mm white cylinder—were used for evaluation. The fine-tuned model accurately predicted angular poses across colours (Fig. 24(A.e-g)) and diameters (Fig. 24(A.h-i)). To further assess generalisation, an untrained pen was tested

(Fig. 24(A.j-k)), also yielding precise alignment. Temporal results (Fig. 24(B)) show that when no contact occurred (Fig. 24(A.a/l)), both predicted angles and forces remained near zero, with minimal terminal noise (Fig. 24(B.a/l)). Upon contact, the robot's orientation (purple line) closely followed the estimated object pose (Fig. 24(B.a-d)), maintaining stable control across colour (Fig. 24(B.e-g)), size (Fig. 24(B.h-i)), and object type (Fig. 24(B.j-k)).

These results confirm that MagicGripper, equipped with fine-tuned mini-MagicTac sensors, can accurately perceive and correct misalignment in real time. The framework demonstrates high generalisation, robustness to object variation, and minimal adaptation data requirements, establishing a foundation for manipulation in unstructured environments.

E. Autonomous Robotic Grasping Task With MagicGripper

To validate the proximity and contact detection algorithms introduced in Fig. 9, an autonomous robotic grasping experiment was conducted based on MagicGripper's multimodal sensing capabilities.

1) Experiment Preparation: All core sensing functions were individually verified before real application:

- **Proximity Detection:**The proximity module must detect approaching objects and differentiate valid proximity cues from illumination noise.As shown in Fig. 25(A), when a pen approached, RGB-channel distributions diverged and overall intensity increased with reduced distance.Correspondingly, channel entropy rose while correlation decreased (Fig. 25(B)), indicating an **"With Proximity"** state (Algorithm 2).Under flashlight interference (Fig. 26(A)), all channels exhibited high similarity, producing strong correlation coefficients despite entropy fluctuation (Fig. 26(B)), classified as **"Light Noise"**.
- **Contact Detection:**Contact detection estimates touch probability without explicitly tracking the grid pattern.In Fig. 27(A), grid geometries vary between channels yet retain high similarity to the reference mask when no contact occurs, consistent with temporal fusion results (Algorithm 1).As MagicGripper closed and the screwdriver touched the surface, grid similarity decreased (Fig. 27(B)), indicating a **"With Contact"** state (Algorithm 3).
- **Combined Detection:**From Fig. 28(A), mixed proximity-contact tests were performed with six everyday objects: a human finger (c), reflective metal spanner (d), red rubber screwdriver (e), green transparent screwdriver (f), white opaque pen (g), and yellow rubber scissors (h).Random light interference was added at the start and end (b, i). In Fig. 28(B), channel entropy fluctuated during object approach or light interference, with amplitude depending on material properties. Diffusely reflective surfaces (e.g., the white pen or yellow scissors) produced stronger entropy variations than specular or transparent ones (e.g., the metal spanner or clear screwdriver), as the latter reflect or transmit light away from the camera. Correlation coefficients decreased steadily during approach but rose sharply to nearly 1.0 under illumination noise. Contact results (Fig. 28(C)) show that all touch events reduced grid similarity from 0.85 to below 0.8, confirming correct contact detection, while illumination noise caused a sharper drop below 0.5.These findings highlight the importance of combining proximity (Algorithm 2) and contact (Algorithm 3) states to enhance robustness under complex lighting.

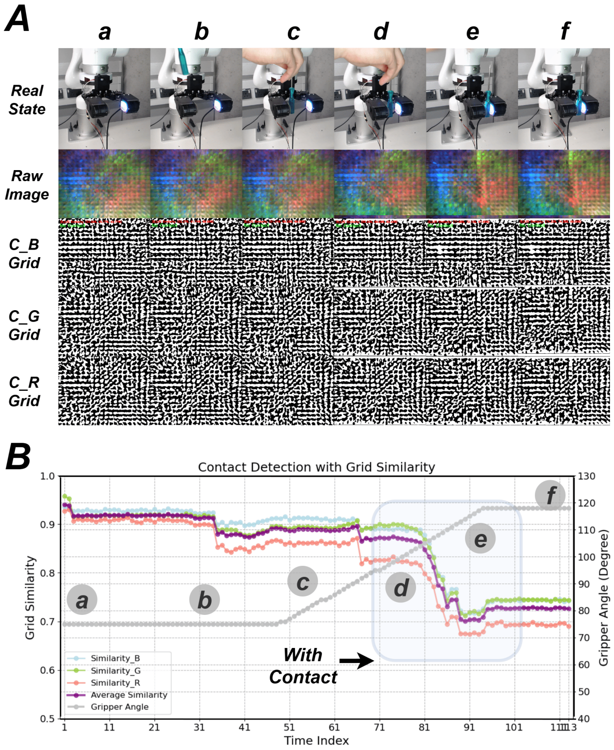


Fig. 27. A: The internal grid deforms only after physical contact occurs. Channel geometries differ (a-c) yet follow a consistent deformation trend upon contact (d-f). B: As MagicGripper touches the screwdriver, grid similarity across channels drops markedly (d-f), indicating a contact state.

2) Experiment Design: Using the validated algorithms, an autonomous grasping task was conducted to demonstrate MagicGripper's fully self-contained multimodal perception. The goal was to enable food-transport manipulation without human supervision, which requires MagicGripper to: (1) detect and localise approaching objects, (2) perform a stable grasp, (3) monitor and respond to slippage, and (4) release the object at a target position. From Fig. 29(A), three everyday food items, an orange, a banana, and a packet of biscuits, were selected to represent varied physical properties.

In details, proximity detection was first activated to identify nearby targets.The proximity probability was computed from differences between total entropy E_{total} and correlation C_{total} relative to thresholds τ_E and τ_C (Algorithm 2).Once an object was consistently classified as "With Proximity", MagicGripper closed gradually by reducing the finger angle.Adaptive thresholding rejected false triggers: a low τ_C might misclassify valid objects as noise, while an excessively high τ_C could overlook weak interference. Upon "With Contact", detection was triggered when grid similarity fell below threshold τ_G, confirming a successful grasp and initiating transport. A sharp increase

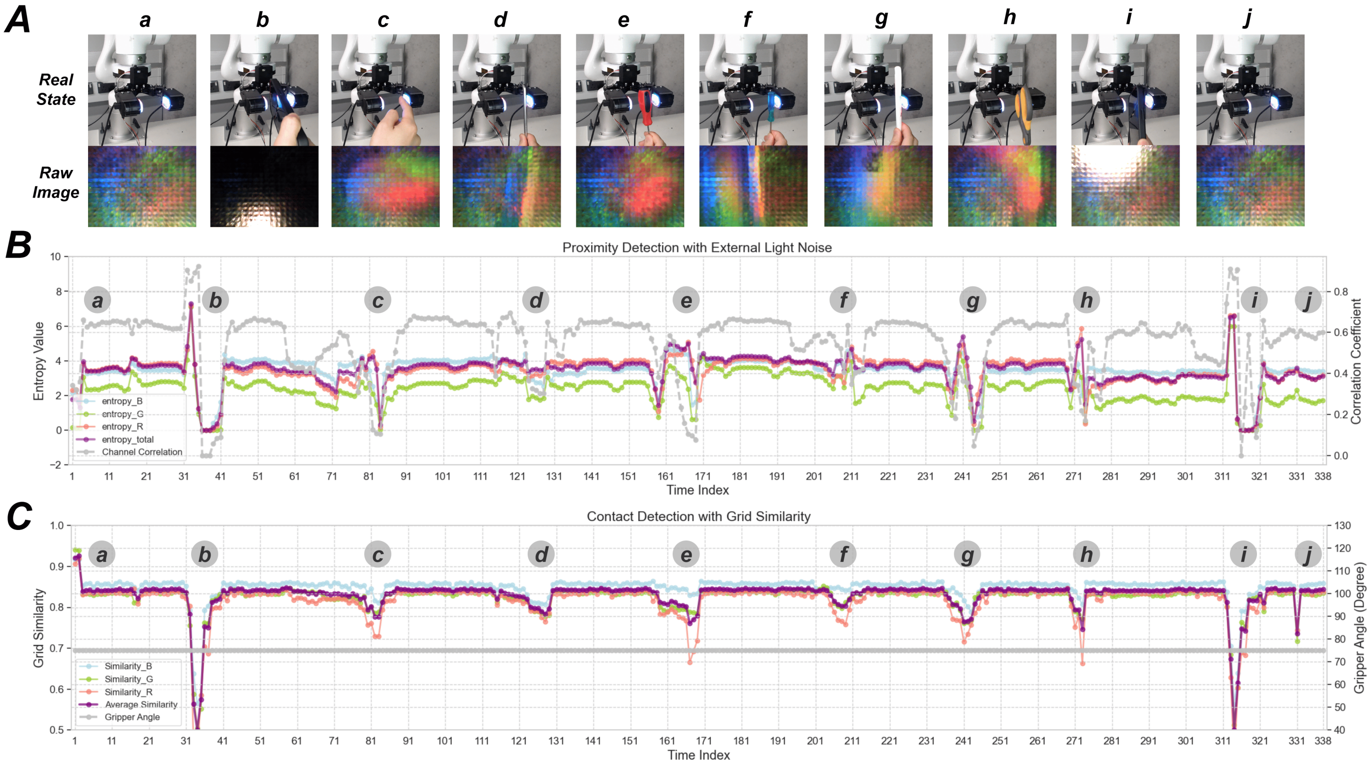


Fig. 28. A: Mixed noise, proximity, and contact tests for MagicGripper with six common objects. B: Distinct effects of light interference and object proximity on channel entropy and correlation. C: Grid similarity changes during light contact and illumination noise.

in similarity during motion indicated slippage, prompting the robot to stop and reset to its initial position.

3) Experiment Analysis: As shown in Fig. 29(B), both proximity and contact detection generalised effectively to the three test items.For the orange (b), banana (e), and biscuits (i), E_{total} increased to around 4 as objects approached, while C_{total} decreased from 0.2 to nearly 0.Even under ambient light interference (h), the features remained separable, confirming algorithmic robustness.

Fig. 29(C) illustrates the correlation between proximity and contact cues.The proximity probability (blue) rose in synchrony with the decline in grid similarity (purple), indicating strong coupling between visual and tactile modalities.With a grid-similarity threshold of 0.7, proximity was typically detected 10-15 frames earlier, providing sufficient lead time for grasp initiation (Fig. 29(D.c/f/j)).

In the three grasp attempts, the first two trials deliberately induced slippage (Fig. 29(A.d/g)), which was successfully detected, causing MagicGripper to stop and return to standby (Fig. 29(C.d/g)). Only the final trial, involving the biscuit pack, completed a full transport and release (k). The operation remained unaffected by light interference (Fig. 29(A.h)), where illumination noise was accurately identified and flagged (yellow trace, Fig. 29(C.h)).

Overall, these results confirm that MagicGripper's multimodal sensing, integrating visual, proximity, and tactile feedback, enables robust, adaptive, and autonomous manipulation of diverse everyday objects. Except for extreme cases such as mirror-like or transparent surfaces, the system achieved reliable autonomy in contact-rich tasks.

V. Discussions

Table VI summarises the main characteristics of the proposed grid-like sensor in multimodal sensing, compared with two representative VBTS types: marker displacement (MD)-based sensors and GelSight-type sensors.

Benefiting from its uniformly distributed multi-layer grid, the proposed sensor exhibits two key advantages. First, the dense grid network extends tactile sensitivity across the entire elastomer, eliminating local "dead zones." Second, its regularly arranged microcells achieve much higher spatial density than sparse marker arrays, enhancing tactile resolution while preserving optical transparency. For example, ViTacTip employs ∼2 mm marker spacing, which limits its ability to capture fine contact geometry compared with GelSight's reflective surface. By contrast, the grid-like structure provides a substantially higher cell density, comparable to GelSight-level resolution while maintaining transparency. This performance is achieved without increasing marker density, which in marker-based designs would otherwise obstruct visual input. Furthermore, the multi-layer grid supports dynamic tactile responses without external force markers or additional calibration layers required by GelSight-type sensors. Integral 3D printing also improves structural consistency and customisability, removing the manual alignment steps typical of conventional fabrication.

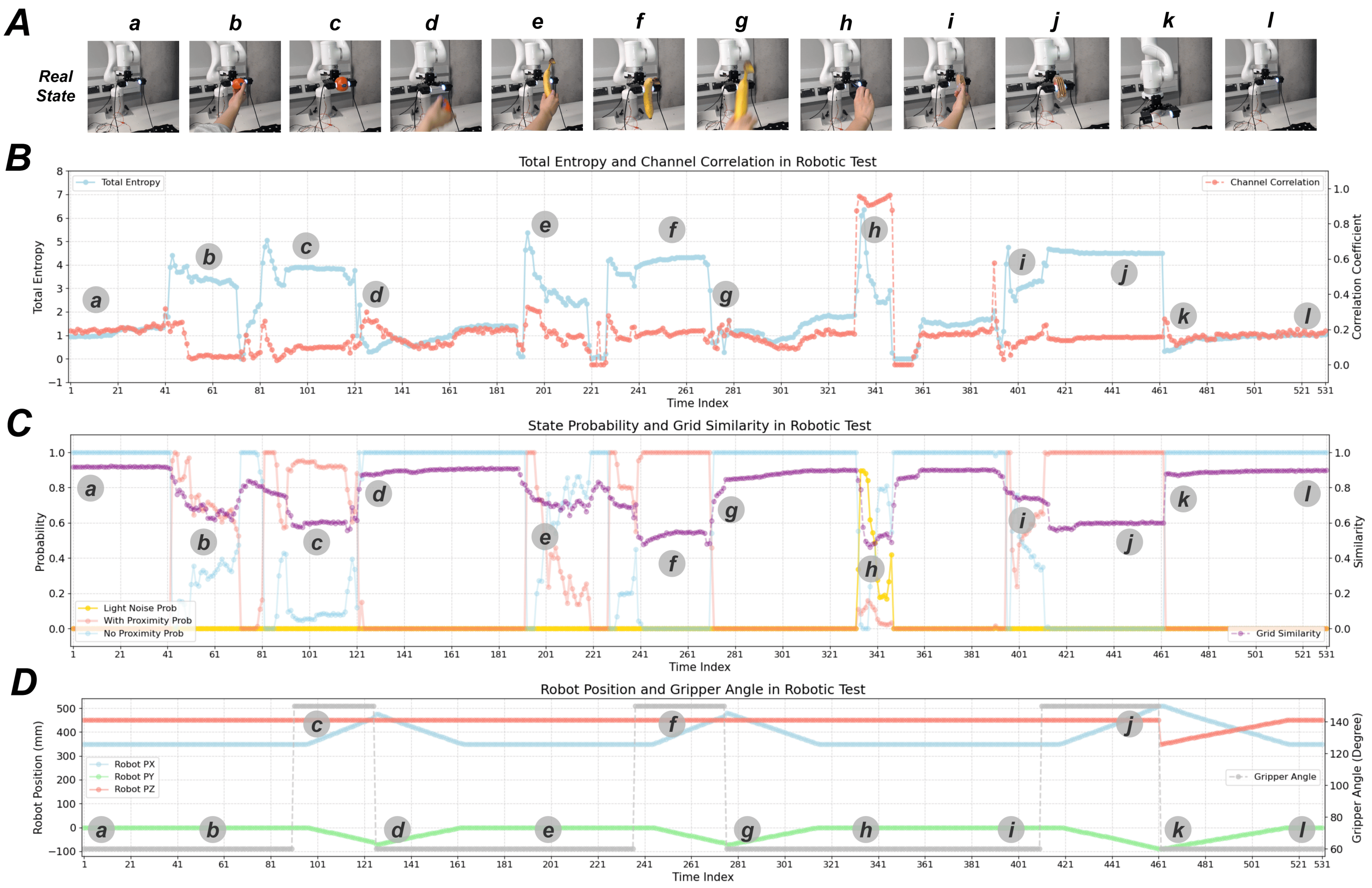


Fig. 29. A: Autonomous grasping experiment with MagicGripper. The system detects approaching objects while rejecting light noise, executes grasping, and transports the object from point A to point B. Upon detecting slippage, the arm automatically returns to the starting position. B: Variations in channel entropy and correlation coefficients distinguish between illumination noise and genuine proximity events. C: Grid similarity serves as a reliable indicator of slippage during transport. D: After two induced slippage events and light disturbances, MagicGripper successfully completes the final grasp-and-transport cycle.

TABLE VI
COMPARISON OF THE PROPOSED GRID-LIKE SENSOR WITH REPRESENTATIVE VBTSS

Sensor Type	Tactile Mechanism	Resolution	Dynamic Response	Sensing Range	Design Flexibility	Fabrication Cost	Multimodal Capability
GelSight-type	Coating Reflection	High	Marker-assisted	Wide	Low	High	Limited
Marker-based	Marker Displacement	Low	Sensitive	Narrow	Low	High	Limited
Grid-like (Ours)	Grid Deformation	Moderate	Sensitive	Wide	High	Moderate	Integrated

Beyond tactile performance, the grid architecture inherently supports multimodal perception by exploiting internal refraction and reflection within grid cells, thereby integrating visual and tactile cues within a single structure. In contrast, marker-based sensors rely on additional components, such as transparent membranes or UV markers [28], [29], to achieve comparable multimodality. These methods introduce trade-offs between tactile resolution and optical clarity, and increase the complexity of modality switching. For instance, enhancing ViTacTip's tactile resolution requires denser opaque markers that block light transmission, while its visual-tactile switching depends on GAN-based models trained on large datasets.

The grid design avoids such conflicts by decoupling optical and tactile functions. Adjusting grid geometry to tune tactile sensitivity has minimal influence on light transmission, ensuring consistent visual feedback. In practice, the grid-like sensor prioritises tactile perception, with vision providing short-range assistance before contact. In MagicGripper, for example, the visual channel activates only within 5 cm of an approaching object, offering reliable pre-contact perception without compromising tactile accuracy. Thus, the modest reduction in transparency does not affect the sensor's overall capability for contact-rich manipulation. Such capability has been further evaluated via a teleoperated assembly task with MagicGripper, as introduced in Supplementary Section VII-B.

VI. CONCLUSION

This paper presented MagicGripper, a multimodal robotic gripper integrating the mini-MagicTac. The proposed design achieves high spatial resolution, sensitive contact detection, and robust multimodal perception without modality interference, while remaining customisable and low-cost through multi-material 3D printing. An accompanying algorithmic framework based on channel entropy and correlation enables reliable proximity and contact detection. Experimental results show that MagicGripper attains a spatial resolution of 0.15 mm, sub-millimetre 3D localisation accuracy, millinewton-level force estimation, and rapid model adaptation using only 5% of new data. Beyond these quantitative results, the core contribution of MagicGripper lies in the seamless integration of visual and tactile modalities within a compact structure, advancing robotic interaction autonomy and offering a scalable, manufacturable pathway toward sensor-rich manipulators capable of contact-aware intelligence.

VII. Supplementary

Theoretical Analysis According to the working principle of the multi-layer grid, we developed a theoretical model for the printed elastomer.

Deformation Analysis: As shown in Fig. 3(A), when the printed elastomer interacts with an external object, the grid cells nearest to the contact point become compressed. The grid structure's constraints and linkages cause a change in the cells' distribution in the surrounding areas, creating a deformation field. In this field, the cuboid-shaped grid cells undergo rotation, translation, squeezing, or stretching, disrupting their parallel alignment. This shift alters the refraction and reflection patterns of incident light. The dimensions of this field are correlated with contact information, such as the contact area, depth, and shape. This correlation enables the mapping of tactile features, including texture, normal and shear force.

Here, a 3D lattice spring model (LSM) has been introduced to describe the deformation property of multi-layer grids within printed elastomer. LSM consists of a particle-based network linked by the spring connection between adjacent nodes. Hooke's and Newton's laws have been widely used to mimic the elastic properties of flexible materials, such as 2D or 3D material. As illustrated in Fig. 2(A), the embedded grid forms a mesh made of Agilus30 and filled core support. For the above structure, we define every eight neighboring cells as a representative elementary volume (REV), whose central point can be used for local elastic analysis, like $P_{m,n,l}$ shown in Fig. 3(B). In this way, a 3 × 3x3 particle-based network can be established, which includes 27 lattice nodes $P_{m+\alpha,n+\beta,l+\gamma}$, where $\alpha,\beta,\gamma \in \{-1,0,1\}$ and $\alpha,\beta,\gamma \neq (0,0,0)$. Each cell has the same size (Δx,Δy,Δz). Here, to simplify the model, the grid cell is set as a cube, i.e. Δx=Δy=Δz=h.

We represent an LSM in an m-dimensional space, where each central node is surrounded by n adjacent nodes, denoted as DmQn. Here, we assume that 1 central node has 14 nodes, which form the D3Q14 framework. Among these nodes, the 6 nearest nodes are defined as structure nodes $P_{structure}$, and the other 8 third-nearest nodes are defined as diagonal nodes $P_{diagonal}$, distinguished in Fig. 3 (B) by green and white circles. Each of them is connected to the central point $P_{m,n,l}$ by a virtual spring, which is painted purple. To define such a spatial relationship between each node position, a shape matrix D_{D3Q14} can be summarised as:

$$D_{D3Q14} = [D_{structure}, D_{diagnal}]_{3\times 14} \tag{4}$$

$D_{structure}$ and D_{diagonal} can be expressed as follows, where the *jth* column is the coordinates of the *jth* neighboring node to the central node located at coordinate (0, 0, 0).

$$D_{structure} = \begin{bmatrix} h & -h & 0 & 0 & 0 & 0 \\ 0 & 0 & h & -h & 0 & 0 \\ 0 & 0 & 0 & 0 & h & -h \end{bmatrix}_{3\times 6} \tag{5}$$

$$D_{\mathrm{diagonal}} = \begin{bmatrix} h & -h & -h & h & -h & h & h & -h \\ h & -h & -h & h & h & -h & -h & h \\ h & -h & h & -h & h & -h & h & -h \end{bmatrix}_{3\times 8} \tag{6}$$

As illustrated in Fig.3(B-b), consider an external force F_d acting on the cube which has a shape matrix D_{D3Q14}, the position of each node should be displaced, and the springs between the nodes are also stretched due to the generated deformation. Here we define the central point $P_{m,n,l}$ with P_i and one of its neighboring nodes as P_j ($j \in [1, N]$ whose order follows D_{D3Q14}, $N = 14$), and their corresponding displacements after deformation are u_i and u_j respectively. In this case, the elastic energy ϕ_i^e which is stored in the central node P_i should be expressed as:

$$\phi_i^e = \frac{1}{2V}\left\{\alpha \sum_{j=1}^{N} g_j\left[(u_j - u_i)\cdot \hat{x}_{ij}\right]^2 + \chi\beta \sum_{ijk} g_j \left[\cos\theta_{ijk} - \frac{\sqrt{2}}{2}\right]^2\right\} \tag{7}$$

In which the first sum part comes from the elasticenergy of all linear springs centered P_i and the second term comes from the total angular springs. $V = h^3$, should be the volume of each cell, α and β are the constants of linear spring and angular spring, $N = 14$ following the order of D_{D3Q14}, $g_j = [g_{structure}, g_{diagonal}] = [2, 3/4]$ is the weight coefficient to ensure local isotropy of the D3Q14, $\hat{x}_{ij} = x_{ij}/\left|x_{ij}\right|$ is the normalized direction vector from P_i to P_j, $\chi = 2/3$ is a constant introduced to keep the isotropy of the coupled spring model, and θ_{ijk} should be the angle between two adjacent linear springs, including P_i to P_j and P_i to P_k.

With the displacement as u_i, then the elasticity generated spring force F_i^e on central node P_i can be calculated by:

$$\begin{aligned} F_i^e &= \frac{\partial \phi_i^e}{\partial u_i} \\ &= \frac{1}{V}\left\{\alpha \sum_{j=1}^{N} g_j\left[(u_j - u_i)\cdot \hat{x}_{ij}\right]\hat{x}_{ij} + \chi\beta \sum_{j=1}^{N} g_j \frac{u_j - u_i}{\left|x_{ij}\right|^2}\right\} \end{aligned} \tag{8}$$

Then the combined force F_i acting on the central node P_i can be expressed as:

$$F_i = F_i^e + F_i^g + F_i^d \tag{9}$$

In which F_i^e is the above calculated spring force, $F_i^g = m_i g$ where $m_i = \rho_i V$ and $\rho_i = \varphi(\rho_{ag}, \rho_{sup})$ which can be defined by print material density of Agilus30 and support, F_i^d is the external force applied on.

After modeling in this way, the combined force F_i applied to P_i is known. Through Newton's second law, we can establish the relationship between the P_i's acceleration a_i and the combined force F_i, and then update the P_i's new positional location at the next moment $t + \Delta(t)$ through discrete time integration. The Verlet algorithm is used to calculate this dynamic particle problem. Given a starting moment of t and an initial condition of the positions u_i, the velocities v_i, and the accelerations a_i of node P_i, then we can calculate v_i using the equations below:

$$u_i(t+\Delta t) = u_i(t) + v_i(t)\Delta t + \frac{a_i(t)}{2}\Delta t^2 \tag{10}$$

$$v_i\left(t+\frac{\Delta t}{2}\right) = v_i(t) + \frac{a_i(t)}{2}\Delta t \tag{11}$$

$$a_i(t+\Delta t) = \frac{F_i(t+\Delta t)}{m_i} - \theta v_i\left(t+\frac{\Delta t}{2}\right) \tag{12}$$

$$v_i(t+\Delta t) = v_i\left(t+\frac{\Delta t}{2}\right) + \frac{a_i(t+\Delta t)}{2}\Delta t \tag{13}$$

The above conclusion can be easily extended to the whole printed elastomer whose size is $X \times Y \times Z$, by letting the dimension of the multi-layer grid be R rows, C columns, and L layers where $R, C, L = \{\lfloor X/\Delta x\rfloor + 1, \lfloor Y/\Delta y\rfloor + 1, \lfloor Z/\Delta z\rfloor + 1\}$ and $R, C, L > 3$. Then there are a total of $(R-1)*(C-1)*(L-1)$ cells and they can be composed of $R*C*L$ lattice nodes, defined as $P_{m,n,l}$ above, where $m, n, l \in \{[0, R], [0, C], [0, L]\}$. Their combined forces and mass can be summarised into matrix format as $F_{R\times C\times L}$ and $M_{R\times C\times L}$. Therefore, Equation 10-Equation 13 could be updated as:

$$U_{R\times C\times L}(t+\Delta t) = U_{R\times C\times L}(t) + V_{R\times C\times L}(t)\Delta t + \frac{A_{R\times C\times L}(t)}{2}\Delta t^2 \tag{14}$$

$$V_{R\times C\times L}\left(t+\frac{\Delta t}{2}\right) = V_{R\times C\times L}(t) + \frac{A_{R\times C\times L}(t)}{2}\Delta t \tag{15}$$

$$A_{R\times C\times L}(t+\Delta t) = \frac{F_{R\times C\times L}(t+\Delta t)}{M_{R\times C\times L}} - \theta V_{R\times C\times L}\left(t+\frac{\Delta t}{2}\right) \tag{16}$$

$$V_{R\times C\times L}(t+\Delta t) = V_{R\times C\times L}\left(t+\frac{\Delta t}{2}\right) + \frac{A_{R\times C\times L}(t+\Delta t)}{2}\Delta t \tag{17}$$

Optical Analysis: As shown in Fig. 4(A), there are several different light propagation modes, including external reflection, internal reflection, and internal refraction, where the last one is the focus of optical modeling. The tactile sensing capability of the printed elastomer relies on internal refraction between grid cells to capture deformation images. Additionally, this process is coupled with the results from two other modes, which complicates optical analysis due to intricate grid architecture of printed elastomer.

Thus, a separate grid cell is extracted for evaluation, the smallest unit inside the REV defined in Fig. 3(B). Assuming that the 8 vertices of this cube-shaped cell are P_i where $i \in [1, 8]$, their corresponding spatial positions can be determined by u_i defined in the previous section. For simplicity purposes, the cell surfaces are all considered to be planar which group a convex surface set $\mathbf{S}$, comprising 6 confined planes $S_{1,2,3,4}$, $S_{5,6,7,8}$, $S_{1,2,5,6}$, $S_{3,4,7,8}$, $S_{1,4,5,8}$ and $S_{2,3,6,7}$. In the following, the whole analysis process is divided into two stages, namely (1) interface propagation and (2) internal propagation.

Interface propagation: This stage occurs on the cell surface and mainly explores how an incident ray $\mathbf{I}_{k-1}$ from the previous phase $k-1$ is transformed into reflection ray $\mathbf{I}_k^l$ and refraction ray $\mathbf{I}_k^r$ at the current phase k. Now suppose that $\mathbf{I}_{k-1}$ is directed from the optical interface $S_{1,4,5,8}$ into the interior of the cell through the incident point P_k. Given that both P_k and $S_{1,4,5,8}$ are known, then the normal N_k can be calculated with the unit normal vector $\mathbf{g}(g_x, g_y, g_z)$. $\mathbf{I}_{k-1}$ can be defined by light intensity I_{k-1}, and light direction $\vec{i}_{k-1}$. Reflection and refraction occur when $\mathbf{I}_{k-1}$ crosses the interface, which lead to the reflection ray $\mathbf{I}_k^l$ and the refraction ray $\mathbf{I}_k^r$:

$$\mathbf{I}_{k-1} = I_{k-1}\cdot\vec{i}_{k-1} \tag{18}$$

$$\mathbf{I}_k^l = I_k^l\cdot\vec{i}_k^l \tag{19}$$

$$\mathbf{I}_k^r = I_k^r\cdot\vec{i}_k^r \tag{20}$$

The light intensities of reflection ray I_k^l and refraction ray I_k^r can be defined by $\mu' I_{k-1}$ and $\mu'' I_{k-1}$, where μ' and μ'' represent reflected intensity factor and refracted intensity factor. $\vec{i}_{k-1}$, $\vec{i}_k^l$ and $\vec{i}_k^r$ are represented by the unit directional vectors of the incident ray $\mathbf{s}(s_x, s_y, s_z)$, unit directional vectors of the reflection ray $\mathbf{s}'(s'_x, s'_y, s'_z)$ and unit directional vectors of the refraction ray $\mathbf{s}''(s''_x, s''_y, s''_z)$. As shown in Fig. 4(A), ε, ε', and ε'' are the angles of incidence, refraction, and reflection respectively. Now define the critical angle at the interface as ε_c, only when the angle of incidence ε is less than ε_c that the refraction ray $\mathbf{I}_k^r$ will be generated. Otherwise, the total internal reflection (TIR) can lead to the only reflection ray $\mathbf{I}_k^l$. The following discussion focuses on the case where $\varepsilon < \varepsilon_c$. According to Snell's law in 3D space:

$$n\sin\varepsilon = n'\sin\varepsilon' \tag{21}$$

$$n(s\times g) = n'(s'\times g) \tag{22}$$

In Equation 21, n and $n'$are the refractive index of the optical medium in front of and behind the interface. Considering that the embedded grid consists of an Agilus30-made mesh and inter-filled support material, thereby let $n = n_{agilus}$ and $n' = n_{support}$ where n_{agilus} and $n_{support}$ represent the refractive index of Agilus30 and support material separately. The vector form of Equation 21 is summarised in Equation 22. Then the light direction of both refraction s' and reflection s'' can be calculated below where (gs) works as the dot product and results to $g_x s_x + g_y s_y + g_z s_z$:

$$s' = \frac{n}{n'}s + g\sqrt{1 - \left(\frac{n}{n'}\right)^2[1-(gs)^2]} - \frac{n}{n'}g(gs) \tag{23}$$

$$s'' = s - 2g(gs) \tag{24}$$

Therefore the reflection ray $\mathbf{I}_k^l$ and the refraction ray $\mathbf{I}_k^r$ can be expressed as:

$$\mathbf{I}_k^l = \mu' I_{k-1}\cdot\left[\frac{n}{n'}s + g\sqrt{1 - \left(\frac{n}{n'}\right)^2[1-(gs)^2]} - \frac{n}{n'}g(gs)\right] \tag{25}$$

$$\mathbf{I}_k^r = \mu'' I_{k-1} \cdot \left[s - 2g(gs)\right] \tag{26}$$

Internal propagation: The second stage occurs mainly within the grid cell and aims to explore how the current refraction ray $\mathbf{I}_k^r$ transfer to the $\mathbf{I}_{k+1}$ when contacting the next interface at phase $k+1$. Suppose that after crossing the current interface (e.g. $S_{1,4,5,8}$), $\mathbf{I}_k^r$ is gradually relayed inside the grid in its direction s' until touching one of the other 5 surfaces (e.g. $S_{5,6,7,8}$) with the intersection point P_{k+1}. Based on Bouguer-Beer-Lambert Law, as light travels through a medium, its intensity decays as the light traveling through it increases, which can be expressed as:

$$I_{k+1} = I_k^r \cdot e^{-\alpha d} \tag{27}$$

where α represents the Napierian absorption coefficient, and d should be the distance that $\mathbf{I}_k^r$ traveled through, denoted by the Euclidean distance between P_k and P_{k+1} as $\left|P_k P_{k+1}\right|$. P_{k+1} can be obtained by using the line $\bar{s}'$ where s' is located to find the intersection with the predefined surface set $\mathbf{S}$. Excluding the special cases of parallel incidence and passing through the vertices, there should be two intersections to find, one of which is $P-k$ and the other is P_{k+1}. Therefore, the final ray $\mathbf{I}_{k+1}$ that arrives at the next interface can be described below, where $\vec{i}_{k+1} = \vec{i}_k = s'$:

$$\mathbf{I}_{k+1} = I_{k+1} \cdot \vec{i}_{k+1} \tag{28}$$

$$\mathbf{I}_{k+1} = I_k^r \cdot e^{-\alpha d} \cdot \left[\frac{n}{n'} s + g\sqrt{1 - \left(\frac{n}{n'}\right)^2 [1 - (gs)^2]} - \frac{n}{n'} g(gs)\right] \tag{29}$$

Through the above two stages of interface propagation and internal propagation, the ejection ray $\mathbf{I}_{k+1}$ can be calculated by giving the incident ray $\mathbf{I}_{k-1}$ and the spatial position of the current cell's angle points P_i. Then the ejection ray $\mathbf{I}_{k+1}$ will be regarded as the incident ray $\mathbf{I}_{k-1}$ again for the next cell until reaching the printed elastomer boundary sized of (X, Y, Z). Based on the above process, iterative calculations are performed for each incident beam to obtain the optical imaging results of the whole printed elastomer.

Teleoperated Assembly Task With MagicGripper Experiment Design: The teleoperated assembly task aims to assist the user in inserting a plug into a small socket using the MagicGripper. This task emulates real-world assembly challenges, emphasizing the importance of tactile feedback in teleoperation. If the plug is misaligned, it collides with the base, producing increased resistance that serves as feedback for corrective action. Conversely, a successful insertion is indicated by minimal resistance and smooth insertion.

The experimental setup is shown in Fig. 30(A). The dark socket baseplate introduces a visual challenge, as users cannot easily confirm insertion under vision-only teleoperation. In the robotic system, the MagicGripper was mounted as the end-effector of a Kinova Gen3 arm. During operation, the gripper's sensor images were streamed to a display, providing visual and tactile feedback to the user. A trial was considered successful when the plug was fully inserted into the socket.

As illustrated in Fig. 5, MagicGripper's multimodal sensing supports this process. The visual modality aids grasp localization and orientation, while proximity feedback helps correct near-edge grasps or sliding. The tactile modality, derived from deformation of the multi-layer grid, captures both static and dynamic contact cues, including excessive shear caused by misalignment.

To further reduce operator workload, a structural similarity index (SSIM) was introduced to quantify the correlation between successive MagicGripper images. SSIM evaluates luminance, contrast, and structural consistency, computed as

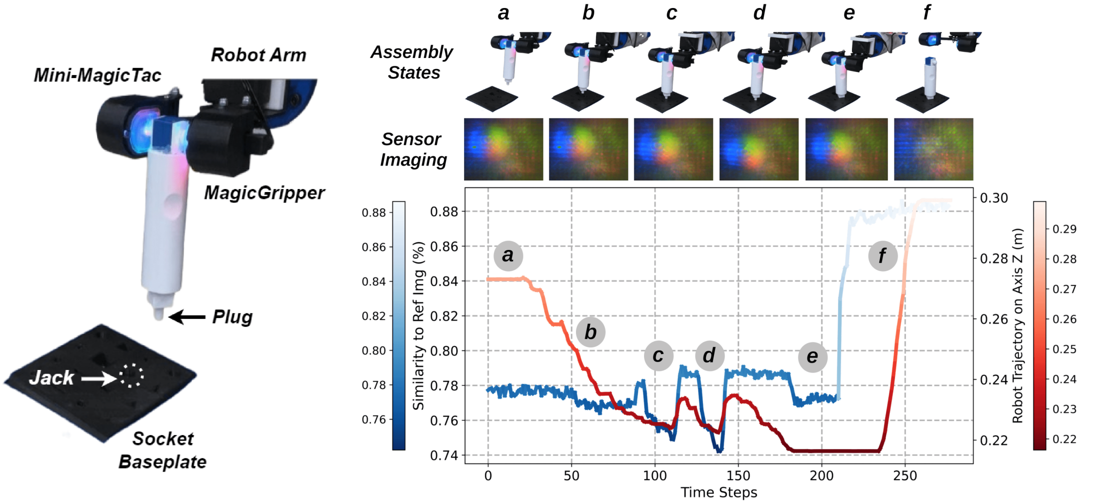
$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{30}$$

where x and y are the compared images; μ_x, μ_y are mean intensities; σ_x^2, σ_y^2 their variances; σ_{xy} the covariance; and C_1, C_2 stabilizing constants. SSIM values range from -1 to 1, with higher values indicating greater similarity.Unlike pixel-wise metrics such as MSE or MAE, SSIM captures perceptually meaningful changes, making it well suited for visual feedback during manipulation. In this study, SSIM compared the current MagicGripper image with a reference (non-contact) state, allowing users to judge insertion success more intuitively, thus improving both efficiency and accuracy.

Results Analysis: The teleoperation process is summarized in Fig. 30(B), where the blue line denotes SSIM variation between consecutive frames and the red line represents the gripper's Z-axis trajectory. The assembly procedure can be divided into six stages:

- **(a) Initial Positioning:** The gripper is positioned diagonally above the baseplate. The initial, contact-free image is used as the SSIM reference.
- **(b) Approach and Alignment:** The user guides the gripper toward the plate while aligning with the socket. The Z-axis decreases gradually. Despite visual occlusion, the high SSIM value (~0.78) indicates no collision.
- **(c) First Insertion Attempt:** A misaligned attempt causes the plug to strike the base, introducing shear deformation in the grid and reducing SSIM to 0.75.
- **(d) Second Insertion Attempt:** After adjustment, another failed insertion occurs with more severe contact, further decreasing SSIM to 0.74.
- **(e) Successful Assembly:** The third attempt achieves proper alignment, reflected by a recovered SSIM of 0.77.
- **(f) End of Task:** Upon successful insertion, the gripper releases and retracts. As the object exits the sensing area, SSIM rises to its peak (0.88).

Sixteen repeated trials were conducted to evaluate the system's ability to detect misalignment, with and without MagicGripper feedback. Controlled angular and lateral offsets exceeding 3 mm were introduced to simulate realistic assembly errors. As summarised in Table VII, MagicGripper achieved a 100% success rate in detecting misalignments, compared with only 25% without tactile feedback. These results highlight the contribution of MagicGripper's multimodal sensing to teleoperation efficiency, allowing the



A. Assembly Task Setup

B. Evaluation Results of Assembly Task

Fig. 30. A: Experimental setup of the teleoperated assembly task, designed to assist the user in accurately inserting a plug into a small jack on the socket baseplate. B: The teleoperated assembly sequence with MagicGripper's multimodal sensing: (a) initial pose, (b) approaching baseplate, (c) first insertion attempt (failed), (d) second attempt (failed), (e) third attempt (succeeded), and (f) end pose.

TABLE VII
ACCURACY OF MISALIGNMENT IDENTIFICATION IN TELEOPERATED ASSEMBLY TASKS

	Misalignment Identification Accuracy
With MagicGripper	100%
Without MagicGripper	25%

operator to quickly identify and correct alignment errors, effectively reducing the reliance on repetitive, vision-only trial-and-error exploration.

REFERENCES

[1] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile sensing for dexterous in-hand manipulation in robotics—A review," *Sensors Actuators A*, vol. 167, no. 2, pp. 171–187, 2011.

[2] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—From humans to humanoids," *IEEE Trans. Robot.*, vol. 26, no. 1, pp. 1–20, Feb. 2010.

[3] F. Yang et al., "Binding touch to everything: Learning unified multimodal tactile representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 26330–26343.

[4] H. Kong, W. Li, Z. Song, and L. Niu, "Recent advances in multimodal sensing integration and decoupling strategies for tactile perception," *Mater. Futures*, vol. 3, no. 2, Jun. 2024, Art. no. 022501.

[5] Q. Mao, Z. Liao, J. Yuan, and R. Zhu, "Multimodal tactile sensing fused with vision for dexterous robotic housekeeping," *Nature Commun.*, vol. 15, no. 1, p. 6871, Aug. 2024.

[6] J. Jiang, X. Zhang, D. F. Gomes, T.-T. Do, and S. Luo, "RoTipBot: Robotic handling of thin and flexible objects using rotatable tactile sensors," *IEEE Trans. Robot.*, vol. 41, pp. 3684–3702, 2025.

[7] S. Liang et al., "AllTact fin ray: A compliant robot gripper with omnidirectional tactile sensing," 2025, *arXiv:2504.18064*.

[8] X. Zhu, B. Huang, and Y. Li, "Touch in the wild: Learning fine-grained manipulation with a portable visuo-tactile gripper," 2025, *arXiv:2507.15062*.

[9] I. H. Taylor, S. Dong, and A. Rodriguez, "GelSlim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 10781–10787.

[10] W. Fan, H. Li, and D. Zhang, "MagicTac: A novel high-resolution 3D multi-layer grid-based tactile sensor," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 388–394.

[11] W. Fan, H. Li, and D. Zhang, "CrystalTac: Vision-based tactile sensor family fabricated via rapid monolithic manufacturing," *Cyborg Bionic Syst.*, vol. 6, 2025, Art. no. 0231.

[12] W. Yuan, S. Dong, and E. Adelson, "GelSight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, Nov. 2017.

[13] D. F. Gomes, Z. Lin, and S. Luo, "GelTip: A finger-shaped optical tactile sensor for robotic manipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 9903–9909.

[14] M. Lambeta 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.*, vol. 5, no. 3, pp. 3838–3845, Jul. 2020.

[15] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "OmniTact: A multi-directional high-resolution touch sensor," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 618–624.

[16] M. Li, T. Li, and Y. Jiang, "Marker displacement method used in vision-based tactile sensors—From 2-D to 3-D: A review," *IEEE Sensors J.*, vol. 23, no. 8, pp. 8042–8059, Apr. 2023.

[17] Y. Yang, X. Wang, Z. Zhou, J. Zeng, and H. Liu, "An enhanced FingerVision for contact spatial surface sensing," *IEEE Sensors J.*, vol. 21, no. 15, pp. 16492–16502, Aug. 2021.

[18] N. F. Lepora, Y. Lin, B. Money-Coomes, and J. Lloyd, "DigiTac: A DIGIT-TacTip hybrid tactile sensor for comparing low-cost high-resolution robot touch," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 9382–9388, Oct. 2022.

[19] K. Sato, K. Kamiyama, N. Kawakami, and S. Tachi, "Finger-shaped GelForce: Sensor for measuring surface traction fields for robotic hand," *IEEE Trans. Haptics*, vol. 3, no. 1, pp. 37–47, Jan. 2010.

[20] X. Lin and M. Wiertlewski, "Sensing the frictional state of a robotic skin via subtractive color mixing," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2386–2392, Jul. 2019.

[21] X. Lin, L. Willemet, A. Bailleul, and M. Wiertlewski, "Curvature sensing with a spherical tactile sensor using the color-interference of a marker array," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 603–609.

[22] W. K. Do, B. Jurewicz, and M. Kennedy, "DenseTact 2.0: Optical tactile sensor for shape and force reconstruction," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2023, pp. 12549–12555.

[23] C. Zhang et al., "GelStereo 2.0: An improved GelStereo sensor with multimedium refractive stereo calibration," *IEEE Trans. Ind. Electron.*, vol. 71, no. 7, pp. 7452–7462, Jul. 2024.

[24] J. Xu, W. Chen, H. Qian, D. Wu, and R. Chen, "ThinTact: Thin vision-based tactile sensor by lensless imaging," *IEEE Trans. Robot.*, vol. 41, pp. 1139–1154, 2025.

[25] W. Kim, W. D. Kim, J.-J. Kim, C.-H. Kim, and J. Kim, "UVtac: Switchable UV marker-based tactile sensing finger for effective force estimation and object localization," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 6036–6043, Jul. 2022.

[26] A. Yamaguchi and C. G. Atkeson, “Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables,” in *Proc. IEEE-RAS 16th Int. Conf. Humanoid Robots (Humanoids)*, Nov. 2016, pp. 1045–1051.

[27] A. Yamaguchi, “FingerVision with whiskers: Light touch detection with vision-based tactile sensors,” in *Proc. 5th IEEE Int. Conf. Robotic Comput. (IRC)*, Nov. 2021, pp. 56–64.

[28] W. Fan, H. Li, W. Si, S. Luo, N. Lepora, and D. Zhang, “ViTacTip: Design and verification of a novel biomimetic physical vision-tactile fusion sensor,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 1056–1062.

[29] Q. Wang, Y. Du, and M. Y. Wang, “SpecTac: A visual-tactile dual-modality sensor using UV illumination,” in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 10844–10850.

[30] F. R. Hogan, M. Jenkin, S. Rezaei-Shoshtari, Y. Girdhar, D. Meger, and G. Dudek, “Seeing through your skin: Recognizing objects with a novel visuotactile sensor,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1218–1227.

[31] S. Athar, G. Patel, Z. Xu, Q. Qiu, and Y. She, “VisTac toward a unified multimodal sensing finger for robotic manipulation,” *IEEE Sensors J.*, vol. 23, no. 20, pp. 25440–25450, Oct. 2023.

[32] S. Zhang et al., “TIRgel: A visuo-tactile sensor with total internal reflection mechanism for external observation and contact detection,” *IEEE Robot. Autom. Lett.*, vol. 8, no. 10, pp. 6307–6314, Oct. 2023.

[33] Z. Song et al., “SATac: A thermoluminescence enabled tactile sensor for concurrent perception of temperature, pressure, and shear,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 5680–5686.

[34] S. Li et al., “M^{3}Tac: A multispectral multimodal visuotactile sensor with beyond-human sensory capabilities,” *IEEE Trans. Robot.*, vol. 40, pp. 4484–4503, 2024.

[35] M. Lambeta et al., “Digitizing touch with an artificial multimodal fingertip,” 2024, *arXiv:2411.02479*.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.