2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Non-contact Tactile Perception in Human-Robot Interaction: Deep Learning-Enhanced Super-resolution Spatial Sensing

Shuyao Zhou[1†], Jikai Liang[1†], Zhengjie Zhu[1], Depeng Kong[3], Zhiao He[1],
Honghao Lyu[2], Geng Yang[1,4*], *Member*, *IEEE*

***Abstract*—With the increasing deployment of robots in dynamic and unpredictable scenarios, it becomes necessary for robots to acquire not only contact-based but also non-contact tactile signals to enhance environmental understanding. However, current non-contact tactile sensors are largely limited to detecting or coarsely recognizing external stimuli, while achieving high spatial resolution typically entails increased sensor density and complex fabrication. This work presents a flexible sparse 2D sensor array, in conjunction with a tailored deep learning model called adaptive spatial-temporal graph convolutional network (ASTGCN), facilitating 3D spatial super-resolution (SR) perception. Built on single-electrode triboelectric nanogenerators with an optimized layout, the sensor array achieves spatial perception while providing a large perception space at low sensor density. Enhanced by the ASTGCN model, this system achieves an average spatial positioning error of 3.11 mm with a physical resolution of only 23 sensors. This research provides novel insights into non-contact haptic perception systems, enabling spatial SR tasks, including spatial trajectory tracking and non-contact gesture classification with 99.33% accuracy, where the gesture classification is used to control a dexterous hand for human-robot interaction.**

I. INTRODUCTION

Robotic perception is crucial for robots to understand the environment and human intentions during manipulation [1], thereby enabling the execution of complex and challenging tasks [2], [3]. Among these, tactile perception, as a rapid and accurate form of sensing, offers a rich and diverse set of information about the environment, particularly in unstructured, dark, or occluded spaces [4], [5]. Various tactile sensors have been proposed based on sensing mechanisms ranging from piezoresistance [6]-[9] and triboelectric [10]-[13] to magnetism [14] and optics [15]. Most tactile sensor signals are generated relying on physical contact, and research has primarily focused on detecting pressure perception [9] and contact position [15] within two-dimensional (2D) planes. As robots are increasingly deployed in dynamic and unpredictable scenarios, contact-based tactile signals are no longer enough for them to timely and accurately perceiving the environment, they are required to acquire non-contact tactile signals as well. These non-contact tactile signals endow robots with three-dimensional (3D) spatial perception, enabling them to more comprehensively understand their surroundings and perform natural human–robot interaction tasks.

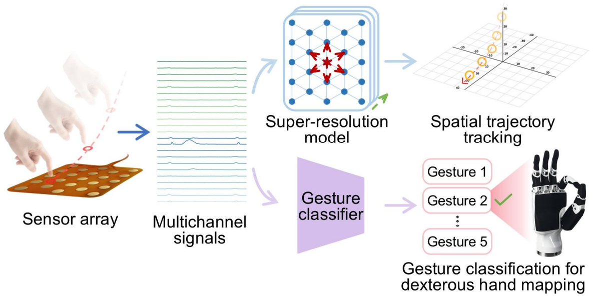

Figure 1. Schematic of the deep learning-enhanced super-resolution spatial tactile perception system.

To address the challenges of 3D spatial perception, researchers have developed tactile sensors based on multiple working mechanisms, including capacitive [16], infrared [17], and magnetic [18]. Despite their effectiveness in spatial perception, the complex fabrication and bulky structures of these sensors restrict their suitability for interactive tasks requiring lightweight sensor structures and flexible designs. In contrast, triboelectric nanogenerator (TENG)-based tactile sensors [19], [20], simple to fabricate, self-powered, and highly sensitive to dynamic stimuli in space, represent a promising solution for non-contact tactile perception. However, most current TENG-based tactile sensors in spatial perception are still restricted to detecting the mere presence of stimuli, with insufficient capacity to provide detailed spatial information, resulting in limited spatial resolution. Despite advances in classical simulation methods for improving perceptual resolution [21]-[24], the gap between simulated and real-world data continues to limit perceptual accuracy.

In this work, we present a 3D spatial super-resolution (SR) tactile perception system based on a flexible and sparse 2D single-electrode triboelectric nanogenerator (SETENG)-based tactile sensor array that can be easily fabricated. This sensor array is designed to enable non-contact tactile perception for human–robot interaction, as shown in Figure 1. An optimized layout of the tactile sensor array, composed of SETENG units sensitive to external spatial stimuli, is proposed to achieve the largest spatial coverage with a minimal number of sensors. Building on this hardware, a tailored model, adaptive spatial–temporal graph convolutional network (ASTGCN), is designed to estimate the 3D spatial position of external stimuli by processing multi-channel signals. Compared with our

---

†Authors contributed equally to this work.

[1]State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, China. [2]Insititute of Advanced Machines, Zhejiang University, Hangzhou, China. [3]Dongfang Electric (Hangzhou) Innovation Institute Co., Ltd, Hangzhou, China. [4]Zhejiang Engineering Research Center of Robotics in Electric Equipment Manufacturing and Intelligent Operation-Maintenance, Hangzhou, China.

*Corresponding author: yanggeng@zju.edu.cn

previous work on contact-based 2D tactile perception [9], [25], the proposed system extends the SETENG array from planar contact sensing to non-contact 3D spatial perception while leveraging deep learning to enhance its perceptual capability. The spatial SR perception system achieved an average spatial positioning error of 3.11 mm (root-mean-square error, RMSE). The system was further applied to challenging tasks, including spatial trajectory tracking along straight lines, figure-eight curves, and spirals, while also achieving 99.33% accuracy in a non-contact gesture classification task with a standard model. These results demonstrate that the proposed spatial SR sensing system enables accurate spatial perception and supports intention-aware human–robot interaction through reliable non-contact gesture classification.

II. RELATED WORK

A. *Non-Contact Sensors Based on Different Working Mechanisms*

Spatial perception has increasingly been recognized as essential for robots engaged in complex task execution. To enable these capabilities, researchers have developed non-contact sensors based on various working mechanisms, including infrared, magnetic, and capacitive [26].

Infrared sensors [27], owing to their rapid response and ease of integration, have been widely employed for non-contact perception, with susceptibility to environmental factors such as ambient light and temperature, limiting reliability in dynamic environments.

Magnetic sensors [28], in contrast, offer more stable performance that is less affected by these factors. However, they typically require magnetic elements and involve complex system designs, which reduces their suitability for flexible human–robot interaction. Capacitive sensors provide an alternative with high sensitivity and low power consumption [29]. However, most of them remain susceptible to environmental disturbances and primarily respond to conductive or high-permittivity materials.

TENG-based tactile sensors have emerged as a promising solution due to their simple fabrication, self-powered, and high sensitivity to dynamic stimuli, enabling effective non-contact spatial perception [30].

B. *TENG-based Tactile Sensors for Robotic Spatial Perception*

TENG-based tactile sensors have been increasingly explored for non-contact and spatial perception tasks. Some studies have utilized a single TENG sensor to detect the approach or withdrawal of external objects [31]. However, these can only detect the presence of external stimuli and provide limited spatial information. To enhance spatial perception ability, some studies have introduced array configurations, allowing for the acquisition of more detailed spatial information.

Zhou et al. [32] developed a non-contact 3 × 3 sensor array utilizing triboelectric and electrostatic induction. The signals generated by the array were processed using a multilayer perceptron (MLP) neural network, enabling the classification of 16 distinct gestures with an average recognition accuracy of 96.5%. In addition, Yang et al. [33] constructed a 4 × 4 array of single-electrode triboelectric nanogenerators (SETENG), which demonstrated the capability for real-time tracking of object motion trajectories. However, the spatial resolution of the system was constrained by the density of its sensing units. Furthermore, Du et al. [34] designed a 20 × 20 TENG-based sensor array, which can achieve a significant enhancement in the perception ability of the 3D shape of non-contact objects by further increasing the density of unit arrangements.

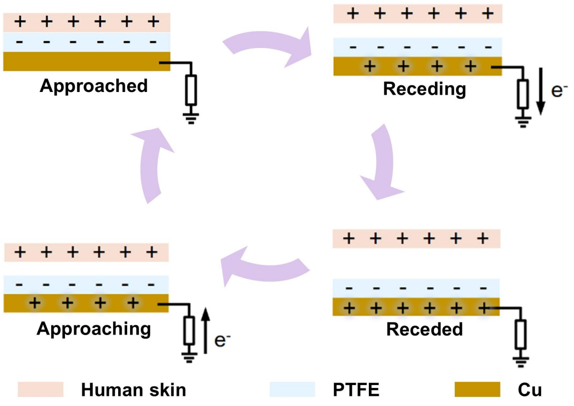

Figure 2. Working mechanism of SETENG-based tactile sensor

Although increasing the density of array configurations can further improve spatial resolution [35], it also increases the difficulty of design and fabrication, restricting their applicability in real-world scenarios. Consequently, how to improve spatial resolution in sparse TENG arrays while accurately capturing dynamic trajectories of external stimuli remains a challenge in spatial perception research.

III. SETENG-BASED TACTILE SENSOR: FABRICATION, MECHANISM, AND PERFORMANCE

SETENG, which is highly sensitive to spatial stimuli, easy to fabricate, and self-powered, represents a promising solution for non-contact tactile perception. It serves as the smallest sensing unit (taxel) in our sensor array, responsible for sensing stimuli in its nearby space.

A. *Working Mechanism and Fabrication Procedure*

The SETENG-based tactile sensor used in this study was fabricated with a polyimide (PI) film as the flexible dielectric substrate and a 5-mm-radius Cu electrode, both produced using commercial flexible printed circuit board (FPCB) processing methods. Then, a polytetrafluoroethylene (PTFE) film tape, cut to the same size as the substrate, was attached on top, serving as the electrification layer that generates charges when approached.

The working mechanism of the SETENG-based tactile sensor is based on the electrostatic induction effect (Figure 2). The electrode is connected in SETENG mode with resistor loads. When an external stimulus (e.g., a bare finger or a silicone sphere) gradually approaches the PTFE film covering the electrode, the difference in electron affinity between the two materials causes negative charges to accumulate on the PTFE surface, while positive charges are carried on the surface of the external stimulus. Consequently, equal and opposite charges are induced on the copper electrode, driving free electrons to flow between the electrode and the external load

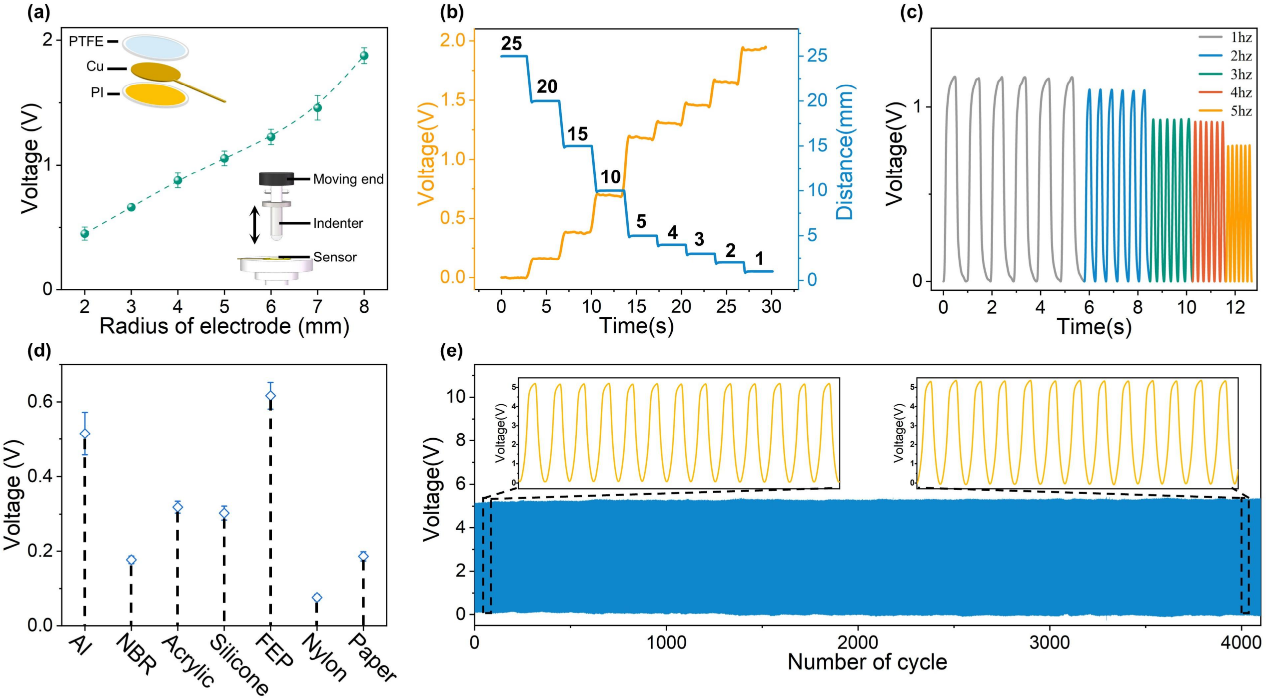

Figure 3. Electrical output of the SETENG-based tactile sensor. The output voltage of SETENG under stimuli of a 25mm diameter silicone hemisphere in (a) different radius of Cu electrode, (b) different approach distances and (c) different approach frequencies. (d)The output voltage induced by various materials close to SETENG including Al, NBR, Acrylic, Silicone, FEP, Nylon and paper. (e) The wear-resisting proposed triboelectric sensor under over 4000 cycles of friction.

to maintain charge balance. As the external stimulus moves closer, additional charges are redistributed between the copper electrode and the load, leading to variations in the output current and voltage. Conversely, when the external stimulus recedes, the induced charges on the electrode are no longer sustained by the external electric field, causing free electrons to flow back and gradually dissipate, thereby restoring the electrical signal to its initial state. In this way, the dynamic process of approach and withdrawal is effectively converted into measurable voltage signals.

B. Electrical Output Characteristics of SETENG

In the characterization experiments, a 3D-printed indenter with an attached silicone hemisphere head (diameter of 25 mm) was connected to the z-axis moving end of a compression testing machine. By controlling the displacement and speed of the z-axis moving end, the indenter approached the sensor to apply external spatial stimuli. The experimental platform setup is shown in the inset of Figure 3a. The output voltage of the sensor, adopted as the characteristic parameter, was measured precisely by a programmable electrometer.

The basic characterization of the tactile sensor was conducted with respect to key parameters that influenced the triboelectric voltage output, such as electrode areas, approach distances, frequencies, and external stimulus materials. For electrode area, the experimental results demonstrate that when the silicone hemisphere approached the sensor at identical distances (from 30 mm to 10 mm), the peak output voltage increased with the electrode radius (Figure 3a). As the approach distance, the output voltage exhibited a stepwise increasing trend as the silicone hemisphere moved from 25 mm to 1 mm toward the sensor. (Figure 3b). Notably, under the same approach distance, an increase in approaching frequency only caused a negligible decrease in the output voltage, which can be almost ignored (Figure 3c). This result indicates that the output characteristics of the SETENG-based tactile sensor are independent of the approaching frequency.

In addition, the sensor generated distinct voltage outputs in response to external stimuli composed of various materials with different electronegativities (Figure 3d), validating its applicability in complex environments. To further evaluate the long-term stability, more than 4000 repetitive contact tests were performed at a frequency of 1 Hz (Figure 3g). The result shows that the output voltage exhibited no noticeable degradation or fluctuation, fully demonstrating the durability and signal repeatability of the sensor.

C. Layout Optimization and Fabrication of the Tactile Sensor Array

Inspired by the principle of SR tactile perception in human skin[36], we propose a SETENG-based tactile sensor array to achieve non-contact tactile perception with high spatial resolution. The working mechanism of SETENG is based on electrostatic induction effect, enabling the perception of external spatial stimuli transmitted through air. With the optimized layout of the array, a single stimulus can simultaneously activate multiple sensors, thereby enhancing spatial localization accuracy. Specifically, when an external stimulus approaches and eventually enters the receptive field of the sensors, it activates adjacent sensors. By synthesizing data from activated sensors, virtual taxels are generated between physical taxels, substantially enhancing the sensor array's spatial resolution.

Overlapped receptive fields can trigger SR perception, it is necessary to investigate the most efficient array layout to

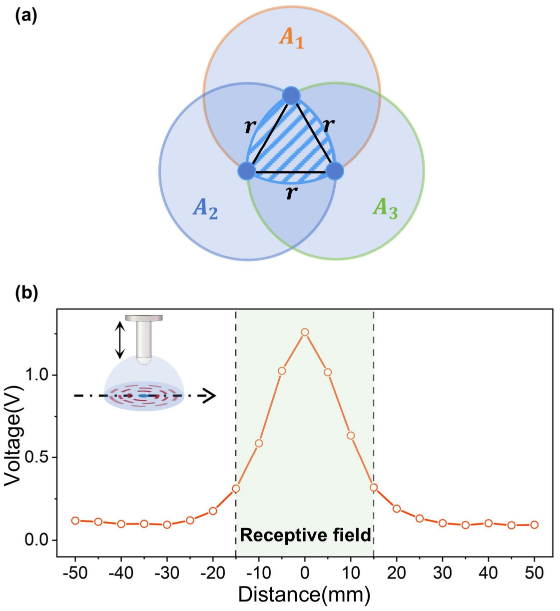

Figure 4. Layout optimization of the tactile sensor array. (a) Optimized taxel layout that maximizes the area of receptive fields. (b)Receptive field test results of a SETENG-based tactile sensor.

maximize the use of available sensing resources. Based on planar geometry, the position of a point in a 2D plane can be uniquely determined from its relative positions to at least three noncollinear reference points[37]. A tactile sensor array only Figure 1. perceives the space above it, so an external stimulus at a certain spatial distance from the array plane (i.e., along the Z-axis) results in distinct responses among the taxels. By analyzing responses from at least three noncollinear taxels, which serve as a basic sensing unit, the location of an external spatial stimulus can be inferred.

The layout of the basic sensing unit follows the optimization method proposed by Kong et al. [9], [25], and the optimization problem can be formulated as follows:

$$\begin{gathered} d_1^*, d_2^*, d_3^* = \arg \max_{d_1, d_2, d_3} (A_1 \cup A_2 \cup A_3), \\ \textit{subject to } d_1 \le r, d_2 \le r, d_3 \le r \end{gathered} \tag{1}$$

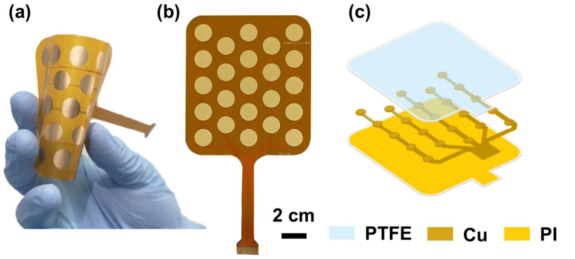

Figure 5. SETENG-based tactile sensor array. (a) Photograph of a fabricated SETENG-based tactile sensor array with high flexibility. (b) The topological structure of the TENG-based tactile sensor array. (c) Exploded view of sensor array.

Where r represents the radius of each taxel's receptive field, $A_1 = A_2 = A_3 = \pi r^2$ is the area of each taxel's receptive field, and d_1, d_2, d_3 denote the distances between taxels. Under the constraint of d_1, d_2, $d_3 \le r$, the positions of the three taxels are optimized to maximize the union of their receptive fields, thereby achieving the largest possible sensing area with the fewest taxels. The optimized layout is obtained when $d_1^* = d_2^* = d_3^* = r$ (as shown in Figure 4a).

Figure 4b illustrates the receptive field test of a single SETENG-based tactile sensor and its output voltage when stimuli were applied at different horizontal distances from the sensor, with each stimulus approaching the sensor along the vertical direction from a height of 30 mm down to 0.5 mm. The results demonstrate a notable disparity in output voltage within 15 mm, which can be considered both the receptive field radius of each taxel and the distance between adjacent taxels.

The optimized layout of the basic sensing unit was extended to construct a flexible tactile sensor array comprising 23 SETENG-based tactile sensors, with a physical size of 80 mm × 72 mm, as shown in Figure 5b. Figure 5c presents an exploded view of this sensor array. To be noted, the fabrication process of the tactile sensor array follows that of a single SETENG-based tactile sensor. This fabrication process, applied uniformly to all sensors, ensures highly consistent electrical properties across the array. As the original physical taxels of the array, these sensors detect the approach and

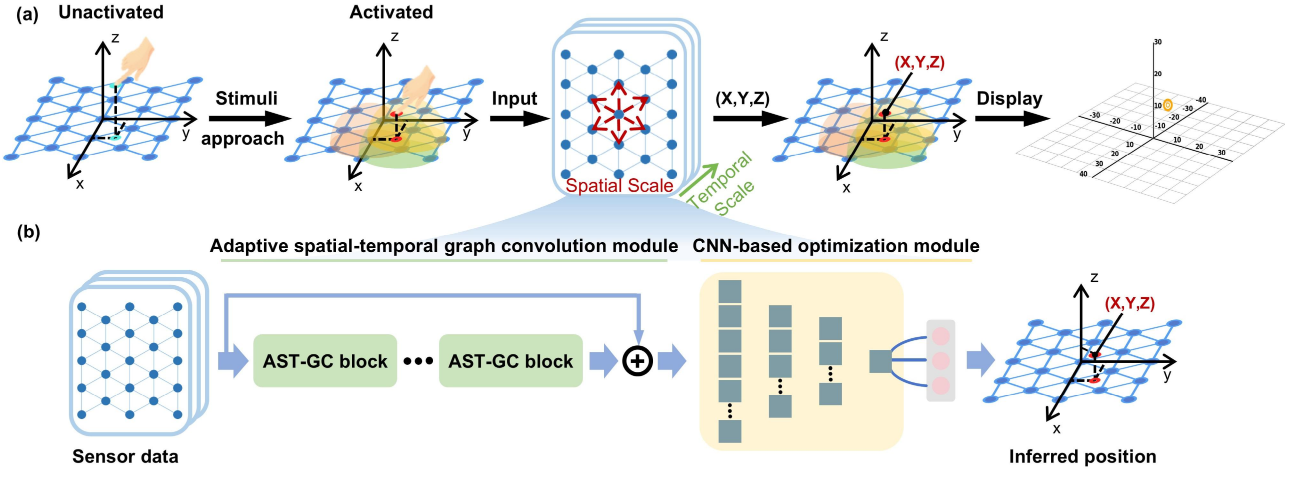

Figure 6. Super-resolution spatial perception method. (a) The overall process of super-resolution spatial perception. (b) Illustration of the overall architecture of ASTGCN, which consists of an AST-GC module and a CNN-based optimization module.

withdrawal of external stimuli through electrostatic induction, producing multi-channel voltage signals that enable non-contact spatial perception.

D. Multi-channel Voltage Acquisition Circuit Design

A simple readout circuit was developed for multi-channel voltage acquisition. Based on the working mechanism of the SETENGs, the potential difference across the resistor to ground of each sensor was used to represent its output voltage, corresponding to the charge induced by external stimuli.

Given the high internal impedance of SETENGs, megaohm-level load resistors were required. For the SETENGs used in this paper, experimental results indicated that 500 MΩ resistors provided an output voltage of suitable amplitude for stable measurements. To ensure high-fidelity, low-noise acquisition, a voltage follower based on an operational amplifier was introduced between the sensor and the analog-to-digital converter (ADC), which transmitted the digitized signals to the controller.

The multi-channel voltage acquisition external board was constructed using three AD7689 ADCs and six TLV274 operational amplifiers. It is equipped with a 24-pin header for connecting the sensor array and is designed to connect to the Arduino Mega 2560 controller via power and communication interfaces. This architecture allows fully synchronous acquisition across all 23 channels, effectively suppressing inter-channel crosstalk and ensuring signal quality.

IV. DEEP LEARNING-ENHANCED SUPER-RESOLUTION SPATIAL PERCEPTION

The SETENG-based tactile sensor array with the optimized layout proposed in the previous section already provides an excellent hardware foundation for SR non-contact tactile perception. Building upon this, deep learning, a powerful technology capable of mapping primary sensory information to functional abstractions, is employed to learn the spatial-temporal features of tactile signals and the hidden connection among sensors within the array, thereby enhancing spatial resolution (Figure 6a).

A. Deep Learning Model for SR Spatial Perception

Due to the remarkable effectiveness of graph neural networks [38] in aggregating information from adjacent nodes, a graph structure is introduced into the inference of external-stimuli spatial location. Accordingly, the layout-optimized tactile sensor array can be conceptualized as a graph, where the nodes represent the tactile sensors and the edges connect adjacent nodes. An adaptive spatial–temporal graph convolutional network, tailored to this structure, is proposed for SR spatial localization of external stimuli above the entire array surface. This model comprises two main components: an adaptive spatial-temporal graph convolution (AST-GC) module and a convolutional neural network (CNN)–based optimization module, as illustrated in Figure 6b.

The AST-GC network [39] is well suited to processing spatial-temporal time-series data and effectively extracting spatial-temporal features from the original sensing signals. Given the distinct differences in response characteristics among sensors in the array, even those manufactured in the same batch, conventional spatial-temporal graph convolutions based on a predefined original adjacency matrix cannot effectively address this issue. The AST-GC module employs a data-driven adaptive adjacency matrix that breaks through the constraints of the predefined adjacency matrix. It captures more diverse relationships among sensors, including non-local and latent dependencies, thereby extracting richer spatiotemporal features and improving the model's performance. Following the AST-GC module, the CNN-based optimization module refines the inferred spatial coordinates, providing more substantial fitting capability and smoother representations, which ultimately improves the accuracy of inferred spatial localization.

A detailed description of these two modules is provided in the following: The AST-GC module consists of nine stacked AST-GC blocks, with the number of output channels for each block set to 8, 8, 16, 16, 16, 32, 32, 64, and 64. Notably, a pruning operation is conducted on this module to reduce model parameters and accelerate computation, making it suitable for the spatial SR localization task. Subsequently, a spatial pooling is performed at the last block to aggregate the spatial features among the N sensors (N=23 in this paper), enabling more effective expression of overall sensing space.

The proposed CNN-based optimization module contains three separate regression branches corresponding to the X, Y, and Z coordinates of the external spatial stimuli. It takes as input the global spatiotemporal features extracted by the AST-GC module. The three outputs are concatenated to form the refined three-dimensional spatial coordinates (X, Y, Z). The structure of each branch, consisting of sequential convolutional layers with output channels of 64, 16, 8, and 1, each followed by a LeakyReLU activation function, is shown in Figure 6b.

B. Spatial Approaching Dataset Acquisition

A Universal Robots UR5 manipulator was used to apply approaching stimuli to the sensor array while recording the spatial coordinates. Attached to the manipulator's end effector, a 3D printed indenter with a hemispherical silicone head (diameter 20 mm) was designed to mimic a human fingertip.

The sensor array (80 mm by 72 mm) was uniformly sampled at 1-mm intervals along both X-axis and Y-axis directions, generating 5913 sampling points in total. For each sampling point, the Z-axis was sampled from 0 mm to 30 mm, representing the stimulus distance to the sensor array plane. While approaching each sampling point to generate the datasets, an Arduino device with a specifically designed multi-channel voltage acquisition external board (Sec. III-D) was used to measure and record the voltage signals, including 23-channel sensor output voltage signals.

The acquired output voltage data were first processed using a low-pass filter to eliminate power frequency noise. These 23-channel time-series signals were then segmented with a sliding window to obtain sequences representing the approach of external stimuli to the array. Min–max normalization was then applied to scale the 23-channel voltage signals to the same range, facilitating comparison of the relative contributions of the sensors. After processing, the data comprised over 100000 samples, which were divided into training, validation, and testing sets with an approximate ratio of 3:1:1, denoted as S_{train}, $S_{validation}$, and S_{test}.

V. EXPERIMENT AND RESULTS

In this section, we validated the effectiveness of the proposed ASTGCN model for SR spatial perception, and applied the co-design of the layout-optimized sensor array and the SR model to challenging tasks, including spatial trajectory tracking and non-contact gesture classification.

A. *Spatial Resolution Enhancement*

The proposed model was first verified for accurately estimating the spatial positions of external stimuli approaching the sensor array based on the original sensor signals. The training set S_{train} was used to determine the hyperparameters of training process, resulting in a batch size of 64 and a learning rate of 0.001. Subsequently, the model was trained on S_{train} to optimize its parameters by minimizing the RMSEs between its outputs and the ground truth. The loss function of out output vector $o^i=(o_1^i, o_2^i, o_3^i)^T$ and its corresponding label $l^i=\left(l_1^i, l_2^i, l_3^i\right)^T$ is shown as follows.

$$L^i\left(o^i, l^i\right)=\sqrt{\frac{1}{3}\sum_{j=1}^{3}\left(o^i - l^i\right)^2} \tag{2}$$

Adam was used with an initial learning rate of 0.001. The learning rate was dynamically reduced by a factor of 10 whenever the training loss did not decrease for 5 consecutive epochs. Training was terminated at epoch 100.

The testing set S_{test} was used to evaluate the trained model, which achieved RMSEs of 3.11, 1.97, 1.77, and 1.64 mm in overall positioning, X-axis positioning, Y-axis positioning, and Z-axis positioning, respectively (Figure 7a). Furthermore, statistics over all testing set S_{test} samples show that more than 80% of the inferred spatial coordinates have relative localization errors within 2.50% (Figure 7b).

B. *Ablation Study*

The ASTGCN was based on the AST-GC network, to which the CNN-based optimization module was added. We used the AST-GC model as the baseline for the ablation study to evaluate the effectiveness of the CNN-based optimization module. Figure 7a shows the evaluation results of the ablation experiments across four evaluation metrics (RMSEs in X-axis positioning, Y-axis positioning, Z-axis positioning and overall positioning). The experimental results show that the proposed model improves the overall spatial localization accuracy by 9.6 % compared with the baseline and achieves effective learning along the three spatial axes (X, Y, and Z).

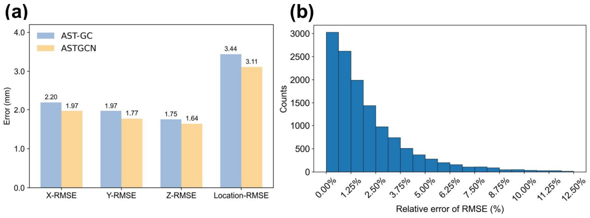

Figure 7. Model evaluation, ablation study and error analysis. (a) Evaluation and ablation results of the model on four metrics. The X-RMSE represents RMSEs in X-axis, the Y- RMSE represents RMSEs in Y-axis, the Z- RMSE represents RMSEs in Z-axis, and the location-RMSE represents RMSEs in overall position. (b)The relative error distribution of the location-RMSE.

Additionally, all experiments were conducted on the PyTorch deep learning framework with the Nvidia GeForce RTX 4090 GPU.

C. *Dynamic Spatial Trajectory Tracking*

Human skin, the largest organ of the body, exhibits high tactile acuity but is generally limited to physical contact perception and cannot effectively sense dynamic spatial stimuli that are non-contact. In this study, the proposed sensor array has surpassed the perceptual limits of human skin which can obtain spatial motion information of external objects before physical contact.

In this experiment, an external object was moved freely above the sensor array along various trajectories, including straight lines (Figure 8a), figure-eight patterns (Figure 8b), and spirals (Figure 8c), while the inferred positions from the model were displayed on the graphical user interface (GUI), highly consistent with the actual spatial trajectories. Notably, the non-contact spatial trajectories inferred in the experiments had never been encountered during model training, yet the model demonstrated high-precision estimation for these unseen trajectories. This indicates that the proposed model possesses strong generalization capability. Furthermore, the co-design of the sparse planar sensor array and the spatial SR algorithm

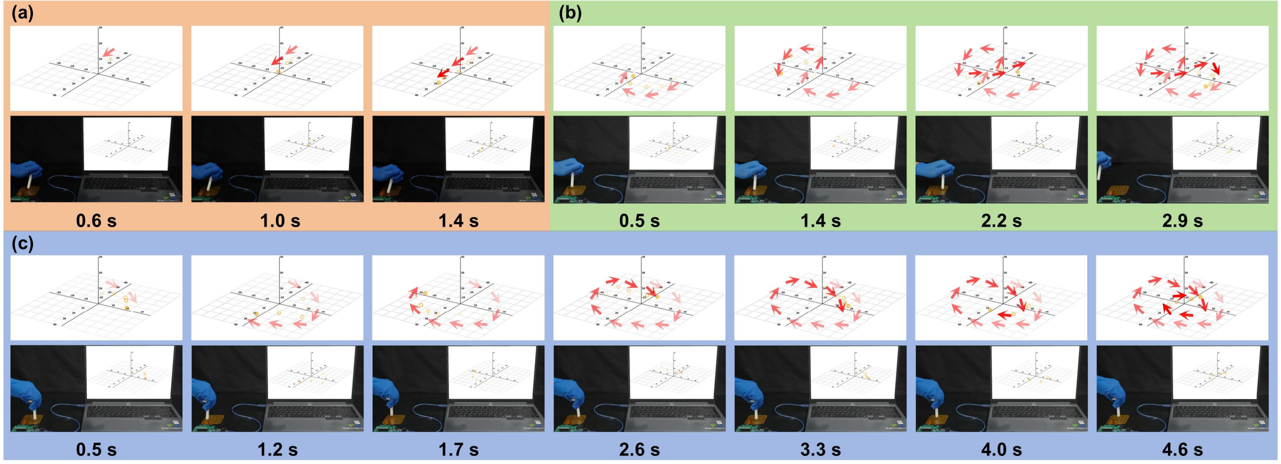

Figure 8. Spatial trajectory tracking experiment: (a) straight line, (b) figure-eight, and (c) spiral, demonstrating good agreement with the actual paths.

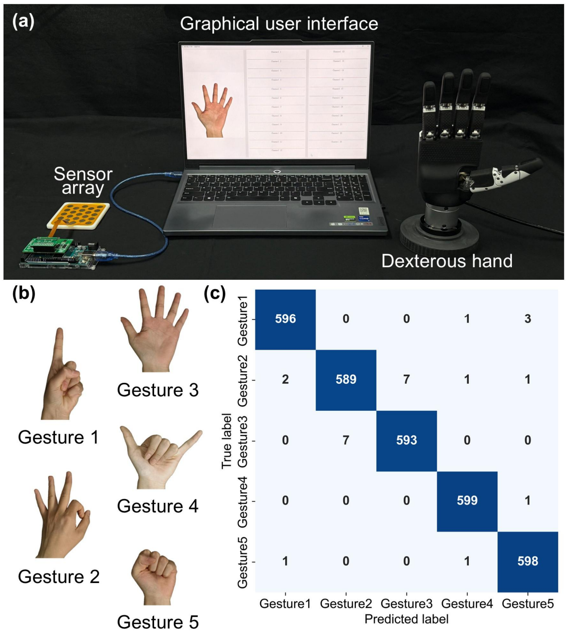

Figure 9. Non-contact gesture classification experiment. (a)Five different gestures. (b)Experimental setup for non-contact gesture classification. (c)Results of the confusion matrix for classifying different gestures.

enable accurate perception of dynamic movements in 3D space, demonstrating its potential for a wide range of non-contact interaction applications.

D. Non-Contact Gesture Classification

Besides spatial trajectory tracking, the SETENG-based tactile sensor array also enables non-contact gesture classification. In human–robot interaction scenarios, the ability to classify hand gestures before physical contact allows robots to anticipate human actions and adjust their behaviors accordingly. This not only enables safer and more natural human–robot interaction but also reduces the limitations and safety risks of traditional contact-based tactile perception.

In the non-contact gesture classification application, data were collected by performing different gestures above the sensor array, as shown in Figure 9b. This process was completed by volunteers to ensure the diversity of data distribution. Approximately 2000 samples were collected for each gesture, with 70% used for training and 30% for validation.

The samples were fed into a CNN-based hand gesture classifier, which employed a 1D convolutional backbone comprising three convolutional layers with output channels of 16, 32, and 32, followed by a fully connected head for classification to infer the current gesture type.

As shown in the confusion matrix (Figure 9c), the CNN-based gesture classifier achieved an accuracy of 99.33% in classifying the five gestures. The classification results were directly mapped to the dexterous hand, enabling real-time replication of gestures (Figure 9a). These results validate the effectiveness of the deep learning-enhanced SETENG-based tactile sensor array for non-contact gesture classification, and demonstrate its capability to enable natural human-robot interaction. Compared with spatial trajectory tracking, non-contact gesture classification further elevates the system from geometric localization to interaction-level understanding, enabling more intuitive and responsive collaboration.

VI. CONCLUSION

In this study, we introduce a novel solution for 3D spatial SR tactile perception through the co-design of a sparse 2D sensor array and a tailored SR model. We designed a flexible and easily fabricated SETENG-based tactile sensor array with an optimized layout, which enables efficient utilization of sensing resources to achieve the largest spatial coverage with a minimal number of sensors. A dataset of over 100,000 approaching stimuli samples was collected for training and validating the proposed SR model, enabling the system to achieve an average spatial positioning error of 3.11 mm. This system was applied to spatial trajectory tracking along straight lines, figure-eight curves, and spirals, as well as to non-contact gesture classification, achieving a classification accuracy of 99.33%.

Despite the promising performance demonstrated in our experiments, several aspects merit further investigation toward practical deployment. The current system primarily considers a single dominant external stimulus, and scenarios involving multiple simultaneous stimuli may introduce signal coupling and increase localization ambiguity. In addition, although preliminary filtering strategies mitigate common power-frequency noise, robustness under stronger electromagnetic interference and dynamically changing operational environments warrants further validation.

The proposed non-contact SR tactile perception system holds strong potential for integration into robotic platforms. The flexible SETENG-based sensor array can be conformally attached to robotic surfaces, such as robotic arms or the inner surface of dexterous hands, enabling continuous perception from non-contact spatial awareness to physical contact. Such integration would allow robots to anticipate object approach, enhance safe human–robot interaction, and achieve more natural intention-aware manipulation. Future work will focus on enabling robust perception under multiple simultaneous stimuli and extending the system to more complex human–robot interaction scenarios.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (No. 52375031, No. 52405038), the Natural Science Foundation of Zhejiang Province (No. LRG25E050001), the Bellwethers +X Research and Development Plan of Zhejiang Province (No. 2024C04057(CSJ), No. 2025C01012, No. 2026C02A1208).

REFERENCES

[1] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, eaat8414, Jun. 2019.

[2] R. Dahiya, "E-Skin: From humanoids to humans," *Proc. IEEE*, vol. 107, no. 2, pp. 247–252, Feb. 2019.

[3] Y. Liu *et al.*, "Aligning cyber space with physical world: A comprehensive survey on embodied AI," *IEEE/ASME Trans. Mechatron.*, pp. 1–22, 2025.

[4] N. Bai *et al.*, “A robotic sensory system with high spatiotemporal resolution for texture recognition,” *Nat. Commun.*, vol. 14, no. 1, 7121, Nov. 2023.

[5] H. Qiao, S. Sun, and P. Wu, “Non-equilibrium-growing aesthetic ionic skin for fingertip-like strain-undisturbed tactile sensation and texture recognition,” *Adv. Mater.*, vol. 35, no. 21, 2300593, May 2023.

[6] H. Lee *et al.*, “Predicting the force map of an ERT-based tactile sensor using simulation and deep networks,” *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 1, pp. 425–439, Jan. 2023.

[7] D. Kong *et al.*, “Bioinspired co-design of tactile sensor and deep learning algorithm for human–robot interaction,” *Adv. Intell. Syst.*, vol. 4, no. 6, 2200050, Jun. 2022.

[8] M. Kim, H. Choi, K.-J. Cho, and S. Jo, “Single to multi: Data-driven high resolution calibration method for piezoresistive sensor array,” *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 4970–4977, Jul. 2021.

[9] D. Kong *et al.*, “Super-resolution tactile sensor arrays with sparse units enabled by deep learning,” *Sci. Adv.*, vol. 11, no. 27, eadv2124, Jul. 2025.

[10] L. Lin *et al.*, “Triboelectric active sensor array for self-powered static and dynamic pressure detection and tactile imaging,” *ACS Nano*, vol. 7, no. 9, pp. 8266–8274, Sep. 2013.

[11] X. Wang *et al.*, “Self-powered high-resolution and pressure-sensitive triboelectric sensor matrix for real-time tactile mapping,” *Adv. Mater.*, vol. 28, no. 15, pp. 2896–2903, Apr. 2016.

[12] Z. Yan *et al.*, “Flexible high-resolution triboelectric sensor array based on patterned laser-induced graphene for self-powered real-time tactile sensing,” *Adv. Funct. Mater.*, vol. 31, no. 23, 2100709, Jun. 2021.

[13] X. Zhi *et al.*, “Hybrid tactile sensor array for pressure sensing and tactile pattern recognition,” *Nano Energy*, vol. 125, 109532, Jun. 2024.

[14] Y. Yan *et al.*, “Soft magnetic skin for super-resolution tactile sensing with force self-decoupling,” *Sci. Robot.*, vol. 6, no. 51, eabc8801, Feb. 2021.

[15] L. Massari *et al.*, “Functional mimicry of Ruffini receptors with fibre Bragg gratings and deep neural networks enables a bio-inspired large-area tactile-sensitive skin,” *Nat. Mach. Intell.*, vol. 4, no. 5, pp. 425–435, May 2022.

[16] Y. Ye, Q. Zhang, X. Li, Y. Liu, Y. Bo, and Y. Lu, “Improving static gesture recognition with capacitive sensor arrays by incorporating distance measurements,” *Measurement*, vol. 257, 118697, Jan. 2026.

[17] Y. Lian *et al.*, “Band alignment semimetal heterojunction-based ultrabroadband photodetector for noncontact gesture interaction with low latency,” *Adv. Mater.*, vol. 37, no. 3, 2404336, Jan. 2025.

[18] P. Makushko *et al.*, “Flexible magnetoreceptor with tunable intrinsic logic for on-skin touchless human-machine interfaces,” *Adv. Funct. Mater.*, vol. 31, no. 25, 2101089, Jun. 2021.

[19] X. Fu *et al.*, “Non-contact triboelectric nanogenerator,” *Adv. Funct. Mater.*, vol. 33, no. 52, 2306749, Dec. 2023.

[20] Y. Li *et al.*, “Advanced dielectric materials for triboelectric nanogenerators: Principles, methods, and applications,” *Adv. Mater.*, vol. 36, no. 52, 2314380, Dec. 2024.

[21] K. Park *et al.*, “A biomimetic elastomeric robot skin using electrical impedance and acoustic tomography for tactile sensing,” *Sci. Robot.*, vol. 7, eabm7187, 2022.

[22] H. Park, K. Park, S. Mo, and J. Kim, “Deep neural network based electrical impedance tomographic sensing methodology for large-area robotic tactile sensing,” *IEEE Trans. Robot.*, vol. 37, pp. 1570–1583, 2021.

[23] H. Sun and G. Martius, “Machine learning for haptics: Inferring multi-contact stimulation from sparse sensor configuration,” *Front. Neurorobot.*, vol. 13, 51, 2019.

[24] B. Wu, Q. Liu, and Q. Zhang, “Tactile pattern super resolution with taxel-based sensors,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2022, pp. 3644–3650.

[25] S. Zhou *et al.*, “Unlocking dynamic subtle stimuli tactile perception: A deep learning-enhanced super-resolution tactile sensor array with rapid response,” *Adv. Intell. Syst.*, vol. 7, no. 5, 2400913, May 2025.

[26] J. Hong *et al.*, “Body-coupled-driven object-oriented natural interactive interface,” *Adv. Mater.*, e07067, Aug. 2025.

[27] R. Wang *et al.*, “Near infrared light-based non-contact sensing system for robotics applications,” *Adv. Mater.*, vol. 37, no. 26, 2414481, Jul. 2025.

[28] J. Ge *et al.*, “A bimodal soft electronic skin for tactile and touchless interaction in real time,” *Nat. Commun.*, vol. 10, no. 1, 4405, Sep. 2019.

[29] S. Mathew and K. Chintagumpala, “A review of recent progress in flexible capacitance pressure sensors: Materials design, printing methods, and applications,” *Adv. Compos. Hybrid Mater.*, vol. 8, no. 3, 236, Jun. 2025.

[30] Y. Tang *et al.*, “Triboelectric touch-free screen sensor for noncontact gesture recognizing,” *Adv. Funct. Mater.*, vol. 30, no. 5, 1907893, Jan. 2020.

[31] C. Zhao *et al.*, “MXene-composite-enabled ultra-long-distance detection and highly sensitive self-powered noncontact triboelectric sensors and their applications in intelligent vehicle perception,” *Adv. Funct. Mater.*, vol. 33, no. 52, 2306381, Dec. 2023.

[32] H. Zhou *et al.*, “Deep-learning-assisted noncontact gesture-recognition system for touchless human-machine interfaces,” *Adv. Funct. Mater.*, vol. 32, no. 49, 2208271, Dec. 2022.

[33] Y. Yang, Y. S. Zhou, H. Zhang, Y. Liu, S. Lee, and Z. L. Wang, “A single-electrode based triboelectric nanogenerator as self-powered tracking system,” *Adv. Mater.*, vol. 25, no. 45, pp. 6594–6601, Dec. 2013.

[34] Y. Du *et al.*, “Multi-receptor skin with highly sensitive tele-perception somatosensory,” *Sci. Adv.*, vol. 10, no. 37, eadp8681, Sep. 2024.

[35] M. Virone, P. Lopes, R. P. Rocha, A. T. De Almeida, and M. Tavakoli, “Dynamic hand gesture recognition using a stretchable multi-layer capacitive array, proximity sensing, and a SVM classifier,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021.

[36] J. M. Loomis, “An investigation of tactile hyperacuity,” *Sensory Processes*, vol. 3, no. 4, pp. 289–302, 1979.

[37] G. Strang, *Introduction to Linear Algebra*. Wellesley, MA, USA: Wellesley-Cambridge Press, 2022.

[38] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A comprehensive survey on graph neural networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021.

[39] L. Shi, Y. Zhang, J. Cheng, and H. Lu, “Two-stream adaptive graph convolutional networks for skeleton-based action recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 12018–12027.