Robotics: Science and Systems 2026
Sydney, Australia, July 13–17, 2026

An Electrical Capacitance Tomography Approach for Robotic Proximity Servoing

Duanpeng Shi[1], Yuliang Wang[1], Yuming Huang[1], Huaping Liu[2*], Di Guo[1*]
[1]College of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
[2]Department of Computer Science and Technology, BNRIST, Tsinghua University, Beijing, China
[*]Corresponding authors: Di Guo (guodi.gd@gmail.com) and Huaping Liu (hpliu@tsinghua.edu.cn)

***Abstract*—Tactile and proximity sensing are fundamental for achieving autonomous robotic manipulation and safe human-robot interaction. However, traditional sensors often face challenges such as environmental interference and the perception gap between far-field vision perception and near-field contact. In this work, we presents a versatile sensing system based on Electrical Capacitance Tomography (ECT) principles, providing a unified framework for non-contact proximity perception, pre-touch orientation estimation and material recognition. We realize the system in two complementary configurations: a large-area array (10 cm $\times$ 10 cm) for high-dynamic safety feedback and a compact module integrated into a robotic gripper (2 cm $\times$ 9 cm). Instead of computationally expensive tomographic reconstruction, we propose CapacitiveServo-Net, a physics-informed neural architecture that extracts spatial dielectric features directly from mutual capacitance measurements. This model facilitates a unified pre-touch servoing framework by mapping high-dimensional capacitive responses to geometric variables (distance and orientation) and material properties. Experimental results on a 7-DOF manipulator demonstrate that our system achieves high-precision, non-contact proximity tracking and real-time pose alignment. Furthermore, the system demonstrates concurrent material classification and pre-touch adaptive grasp refinement during the approach phase, offering a unified solution for proactive perception and manipulation in occluded or degraded visual environments.**

***Index Terms*—electrical capacitance tomography, proximity sensing, servoing framework.**

I. Introduction

As robotic systems move from structured industrial settings into unstructured, human-centered environments, achieving both safety and dexterity in physical human–robot interaction (pHRI) has become a major challenge for large-scale deployment[15]. In collaborative scenarios, robots must move beyond passive safety mechanisms that only mitigate impact after contact occurs. Instead, they require proactive sensing capabilities to perceive nearby objects, anticipate potential risks, and react before physical interaction takes place[16, 17].

Beyond safety, anticipatory perception is also critical for robotic manipulation, where end-effectors must operate reliably under visual occlusion, clutter, and adverse environmental conditions[1].

Despite significant advances in robotic perception, a pronounced gap still exists between far-field vision and near-field contact sensing. Vision-based systems provide strong global scene understanding and and support high-level motion planning. However, their performance often deteriorates in the

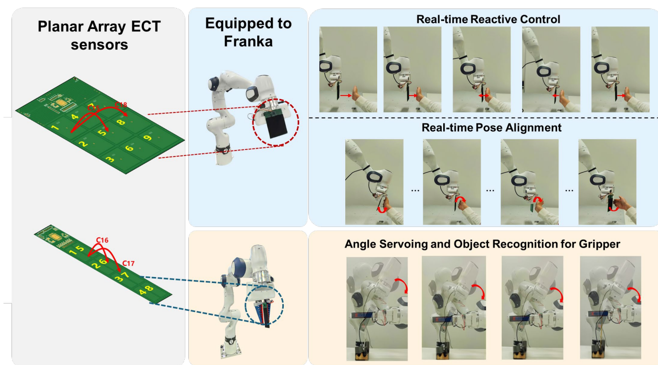

Fig. 1: Overview of the Planar Array Electrical Capacitance Tomography (ECT) Sensing System.

near-contact interaction zone (approximately 0–20 cm from the robot surface) due to self-occlusion, lighting variations, and limited depth-of-field[8]. In contrast, tactile sensors offer rich and localized feedback during physical interaction. Yet, they remain inherently reactive: meaningful information is only available after contact occurs, which may be too late to prevent damage to fragile objects or guarantee safety during high-speed motion[14]. To bridge this perceptual gap, electronic skins based on electric-field sensing have emerged as a promising solution for near-field perception. Among these approaches, capacitive sensing is particularly attractive due to its high bandwidth, omnidirectional sensitivity, and natural responsiveness to biological tissues with high permittivity[5]. However, most existing capacitive skins operate as collections of discrete sensing elements or simple proximity triggers[4]. Consequently, they primarily function as virtual bumpers that can detect the presence of nearby objects, but cannot resolve spatial distribution, geometry, or motion. This limited spatial expressiveness restricts their applicability in closed-loop servoing and dexterous manipulation tasks.

To address these limitations, we present a unified hardware–software framework based on Electrical Capacitance Tomography (ECT) principles for multi-scale perception and control. As shown in Fig. 1, the proposed system supports both square and strip sensor configurations, enabling seamless integration with robotic manipulators and grippers. Unlike conventional capacitive skins that operate as discrete proximity triggers, our approach exploits high-dimensional mutual-capacitance fields to achieve spatially continuous perception. We further introduce a direct perception-to-action pipeline based on a physics-informed neural architecture. Instead of performing computationally expensive tomographic reconstruction, the proposed model extracts geometric primitives directly from raw capacitance perturbations. This design enables proactive pre-touch perception and provides a continuous sensing modality between far-field vision and physical contact. The primary contributions of this work are threefold:

- **Versatile ECT Sensing System:** We develop an open-boundary Electrical Capacitance Tomography (ECT) system optimized for robotics integration. The system includes: (1) a large-area planar "Safety Skin" ($10, \mathrm{cm} \times 10, \mathrm{cm}$) integrated with a high-bandwidth control loop for reactive collision avoidance and angle alignment; and (2) compact strip sensors ($2, \mathrm{cm} \times 9, \mathrm{cm}$) embedded in a robotic gripper for non-contact pre-touch perception and material-aware manipulation.
- **Physics-Informed Perception via CapacitiveServo-Net:** We propose *CapacitiveServo-Net*, a lightweight neural architecture that directly maps mutual-capacitance perturbations to geometric and semantic features. By incorporating a physics-informed target transformation, the framework compensates for the inherent nonlinearity of capacitive sensing. This formulation enables stable training and constant-time inference, making the framework well suited for high-frequency closed-loop control.
- **Closed-Loop Validation in Pre-touch Manipulation:** We validate the proposed framework on a 7-DOF robotic manipulator through real-time proximity servoing and pose-alignment tasks. The system further enables concurrent material recognition and adaptive grasp refinement during the approach phase, extending robotic perception capabilities in visually degraded and occluded environments.

II. Related Work

A. Near-Field Sensing for Safe Interaction and Pre-touch Manipulation

In physical human–robot interaction (pHRI), the ISO/TS 15066 standard formalizes the Power and Force Limiting (PFL) paradigm, which constrains robot motion and interaction forces to ensure human safety[2]. Although effective, such constraints often impose conservative speed limits and reduce operational efficiency. To mitigate this trade-off, prior work has explored dynamic safety strategies based on proactive perception, allowing robots to adapt their behavior according to the proximity of humans or surrounding objects.

Conventional approaches typically rely on external vision systems or LiDAR[12]. While these sensors provide broad scene-level information, they are susceptible to occlusion and often struggle to monitor the entire robot body during close interaction. In contrast, on-board electronic skins provide distributed sensing directly on the robot surface. Resistive or force-based skins respond only after contact, whereas proximity-sensitive skins enable pre-collision reactions, allowing robots to maintain safety margins while operating at higher speeds[7, 20].

Beyond safety, near-field perception is also critical for manipulation, where end-effectors must infer object properties and interaction cues immediately before physical contact. Optical proximity sensors, such as infrared sensors, are inexpensive and easy to deploy, but their measurements are sensitive to surface reflectivity and ambient illumination[10]. Acoustic approaches, including ultrasound, can extend the sensing range but typically suffer from limited spatial resolution, dead zones, and relatively slow response times[9]. Electric-field and capacitive sensing have therefore attracted increasing attention for pre-touch perception, owing to their robustness, high responsiveness, and sensitivity to material properties[7].

Early work on electric-field servoing showed that field gradients can guide gripper motion toward nearby objects[18]. More recently, data-driven methods have enabled multidimensional capacitive servoing, for example in estimating human limb poses from capacitive measurements[5]. However, these methods typically regress raw capacitive signals directly to task variables, without explicitly modeling the spatial structure of the sensed environment. This limits their ability to generalize to novel object geometries, poses, and interaction scenarios.

B. Electrical Capacitance Tomography in Robotics

Electrical Capacitance Tomography (ECT) has been extensively studied in industrial process monitoring, such as multiphase flow estimation in pipelines[21]. Translating ECT to robotic perception, however, introduces two key challenges: open-boundary sensing conditions and stringent real-time constraints. Prior robotic applications have integrated ECT into grippers for internal object inspection, such as liquid-level estimation[13]. However, these systems typically operate in static or quasi-static settings and are not designed for closed-loop interaction. Recent work has further explored ECT for external perception tasks, including non-destructive testing[3]. Our work differs from these studies by using ECT feedback for high-bandwidth closed-loop control, enabling dynamic non-contact interaction and pose estimation without relying on exploratory touch.

The proposed ECT-based pose-servoing framework is closely related to image-based visual servoing (IBVS), where control is driven by the relationship between visual features and robot motion[6]. In contrast to conventional ECT pipelines that rely on computationally expensive tomographic reconstruction, our framework uses *CapacitiveServo-Net* to extract latent dielectric features directly from raw mutual-capacitance perturbations. This direct mapping from capacitive field responses to geometric primitives, such as distance and orientation, enables high-bandwidth pre-touch servoing that remains robust to lighting variations and visual occlusions.

These properties make ECT a promising near-field sensing layer for bridging far-field vision and tactile interaction.

III. Sensor Design, Fabrication, and Characterization

We develop two distinct Electrical Capacitance Tomography (ECT) sensor configurations: a large-scale flexible Safety Skin and a compact Smart Gripper. Together, they support wide-area proximity monitoring and high-resolution pre-touch perception for manipulation.

A. Theoretical Preliminary

Both terminals utilize Electrical Capacitance Tomography (ECT) in mutual capacitance mode to detect proximity and contact. As illustrated in Fig. 2, the system sequentially excites an electrode while measuring induced currents at others. Theoretically, the mutual capacitance C_{ij} between an electrode pair (i, j) is determined by the dielectric distribution $\epsilon(\mathbf{r})$ within the sensing volume Ω:

$$C_{ij} = \frac{1}{V_i V_j} \int_{\Omega} \epsilon(\mathbf{r}) \mathbf{E}_i(\mathbf{r}) \cdot \mathbf{E}_j(\mathbf{r}) d\Omega \tag{1}$$

where V_i, V_j are the excitation voltages and $\mathbf{E}_i(\mathbf{r}), \mathbf{E}_j(\mathbf{r})$ represent the electric field distributions when electrodes i and j are independently excited. When an object (e.g., a human limb or a target for manipulation) enters the domain, it perturbs the fringing electric fields (shown as red lines in Fig. 2), leading to a change in the dielectric constant $\Delta\epsilon(\mathbf{r})$ and a subsequent variation in mutual capacitance ΔC_{ij}. Here, $\mathbf{c}^{(k)} = [\Delta C_{12}^{(k)}, \Delta C_{13}^{(k)}, \ldots, \Delta C_{(M-1)M}^{(k)}]^\top \in \mathbb{R}^{M(M-1)/2}$ denotes the capacitance feature vector of the k-th sample, where each element corresponds to the mutual capacitance variation of an independent electrode pair. M represents the number of electrodes.

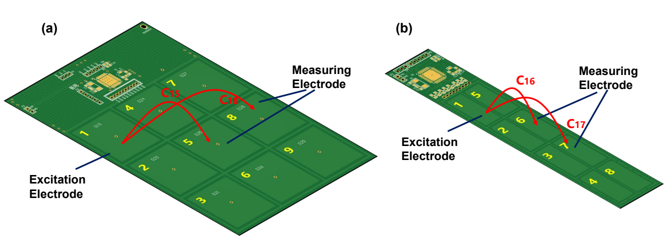

Fig. 2: The fabrication and principle of ECT sensors with two different configurations.

B. ECT Sensor with Two Configurations

a) Configuration 1: Safety Skin (3×3): For wide-area collision avoidance and pose alignment, we design a flexible array comprising $M = 9$ square electrodes arranged in a 3×3 matrix (100 $\times$ 105 mm). Each electrode (28 $\times$ 28 mm, 32 mm pitch) provides a symmetric response across horizontal, vertical, and diagonal axes. This configuration generates a feature vector $\mathbf{c}_{skin} \in \mathbb{R}^{36}$.

b) Configuration 2: Smart Gripper (2×4): To support fine manipulation for robotics, a rigid sensor strip is developed with $M = 8$ rectangular electrodes (20 $\times$ 9.5 mm, 11.5 mm pitch) in a 2×4 array. This layout is specifically optimized for the high-aspect-ratio geometry of parallel grippers and produces a feature vector $\mathbf{c}_{grip} \in \mathbb{R}^{28}$.

C. Noise Mitigation and Signal Characterization

To improve signal integrity for proximity and pre-touch perception, both sensor configurations adopt a shielded PCB architecture. Grounded copper planes are incorporated to provide electromagnetic shielding and a stable reference potential, while the onboard electronics are integrated close to the sensing electrodes. These electronics include the capacitance-to-digital converter (CDC), filtering components, and communication interfaces. This compact integration reduces parasitic lead capacitance and suppresses external interference, thereby improving the signal-to-noise ratio (SNR) for robust near-field sensing.

We further characterize the $2\,\text{cm} \times 9\,\text{cm}$ strip sensor to evaluate its suitability for closed-loop control and pre-touch manipulation. Signal integrity is assessed across all 28 mutual-capacitance channels. As shown in Fig. 3(a), (b), and (e), all channels exhibit stable baselines, with the maximum drift remaining below the 0.5% threshold and low statistical variance across measurements. The inter-channel correlation matrix in Fig. 3(c) shows limited crosstalk, indicating that the sensor preserves spatially discriminative capacitive responses. In addition, the spectral analysis of the most noise-sensitive channel in Fig. 3(d) reveals a broadband noise profile without dominant periodic interference. This enables high-bandwidth closed-loop control with minimal filtering latency.

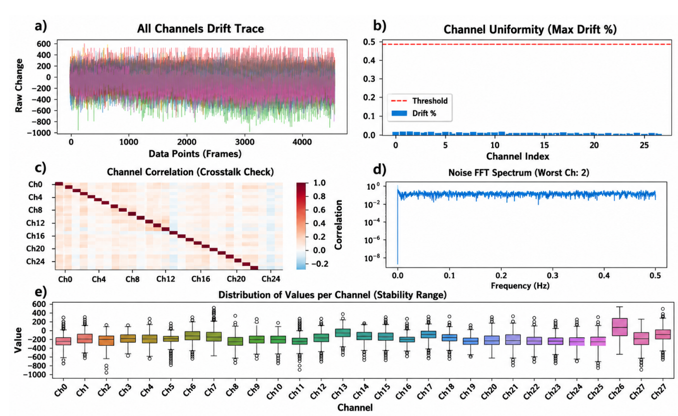

Fig. 3: Hardware performance and signal characterization of the $2\,\text{cm} \times 9\,\text{cm}$ ECT strip sensor.

To evaluate repeatability under practical operating conditions, we also test the sensor array over five independent power-cycle sessions. The observed cross-session consistency indicates that the measured capacitive responses primarily reflect object-induced dielectric perturbations rather than hardware-induced artifacts. This repeatability is important for long-term robotic applications, such as autonomous grasping and pHRI, where the sensing model should maintain reliable perception without session-specific recalibration. Additional characterization results for both sensor configurations are provided in the Appendix.

IV. Methodology

We propose a unified capacitive servoing framework that maps mutual-capacitance perturbations to geometric and semantic variables for pre-touch robotic interaction. Given a temporal window of baseline-subtracted capacitance measurements, the perception module estimates task-relevant quantities, including proximity, orientation, and material class. These outputs are then used by a closed-loop controller for pose alignment, proximity-based velocity modulation, and material-aware grasp refinement.

A. Physics-Informed Perception: CapacitiveServo-Net

Capacitive measurements are high-dimensional and can be affected by electromagnetic interference (EMI), parasitic coupling, and environmental variations. To enable real-time pre-touch servoing on edge computing devices, we introduce *CapacitiveServo-Net*, a lightweight spatio-temporal architecture that maps mutual-capacitance perturbations to task-level geometric and semantic variables. An overview of the framework is shown in Fig. 4.

1) Physics-Informed Target Transformation: A key challenge in capacitive proximity sensing is the strongly nonlinear relationship between capacitance and object distance. In general, the capacitive response increases rapidly in the near field and decays quickly as the object moves away, which creates an imbalanced regression problem across the sensing range. Directly regressing the Euclidean distance d therefore tends to emphasize far-field errors while reducing sensitivity in the near-contact region, where accurate control is most critical.

To address this issue, we introduce a physics-informed target transformation. Instead of predicting distance directly, the network estimates a normalized *Proximity Index* $\rho \in [0, 1]$:

$$\rho = \mathcal{T}(d) = \frac{\delta}{d + \delta}, \tag{2}$$

where δ is a regularization constant that controls the scale of the near-field response. This bounded inverse mapping reflects the distance-dependent decay of capacitive coupling and compresses the dynamic range of the regression target. As a result, it improves numerical conditioning and encourages the network to preserve sensitivity in the pre-touch region.

During inference, the predicted proximity index $\hat{\rho}$ can be converted back to physical distance using the inverse transformation:

$$\hat{d} = \mathcal{T}^{-1}(\hat{\rho}) = \delta \left(\frac{1}{\hat{\rho}} - 1 \right). \tag{3}$$

The same proximity representation is also used directly by the downstream controller for velocity modulation and safety-field generation.

2) Shared Physics-Aware Backbone: The backbone takes a temporal window of baseline-subtracted mutual-capacitance measurements, $\mathbf{X} \in \mathbb{R}^{B \times T \times C}$, where B is the batch size, T is the temporal window length, and C is the number of independent electrode pairs. To capture temporal variations in the capacitive response while remaining lightweight for real-time inference, we use a Multi-Scale Temporal CNN (MS-TCN). Specifically, three parallel one-dimensional convolutional branches with kernel sizes $k = \{3, 5, 9\}$ extract features at different temporal scales, allowing the model to capture both smooth motion-induced trends and short-term signal fluctuations.

Capacitive perturbations are also spatially structured across electrode pairs. When the sensor approaches an object with an off-normal pose, strong responses are typically concentrated in a subset of channels corresponding to the most strongly coupled electrode pairs. To exploit this structure, we incorporate a Squeeze-and-Excitation (SE) block [11] after the temporal feature extraction stage. The SE block adaptively reweights feature channels, allowing the network to emphasize informative electrode-pair responses while reducing the influence of weak or noise-dominated channels. The resulting latent representation is denoted as $\mathbf{z}$.

3) Task Heads: Given the shared latent representation $\mathbf{z}$, CapacitiveServo-Net uses task-specific decoding heads for different deployment scenarios. The heads share the same backbone but predict different quantities required by safety monitoring, pose servoing, and material-aware manipulation.

- **Proximity Head.** This head predicts the proximity index $\hat{\rho}$ for near-field distance estimation and collision avoidance. It is trained with the Log-Cosh loss [19],

$$\mathcal{L}_{prox} = \sum_i \log\left(\cosh(\hat{\rho}_i - \rho_i)\right), \tag{4}$$

  which behaves similarly to an L_2 loss for small errors while being less sensitive to large outliers.
- **Alignment Head.** This head estimates the spatial misalignment $\hat{\boldsymbol{\theta}} \in \mathbb{R}^k$ between the sensor and the target. We use $k = 2$ for skin-based normal alignment and $k = 1$ for gripper yaw alignment. The head is trained with the mean-squared error loss,

$$\mathcal{L}_{align} = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2. \tag{5}$$

  This output captures capacitive asymmetry across the electrode topology and is used by the controller for pre-touch pose correction.
- **Property Head.** This head predicts object material categories for material-aware manipulation. It uses the temporal capacitive response encoded in $\mathbf{z}$ to estimate a categorical distribution $\mathbf{p} \in \mathbb{R}^K$ over K material classes. The head is trained using the cross-entropy loss,

$$\mathcal{L}_{cls} = -\sum_{k=1}^{K} y_k \log p_k. \tag{6}$$

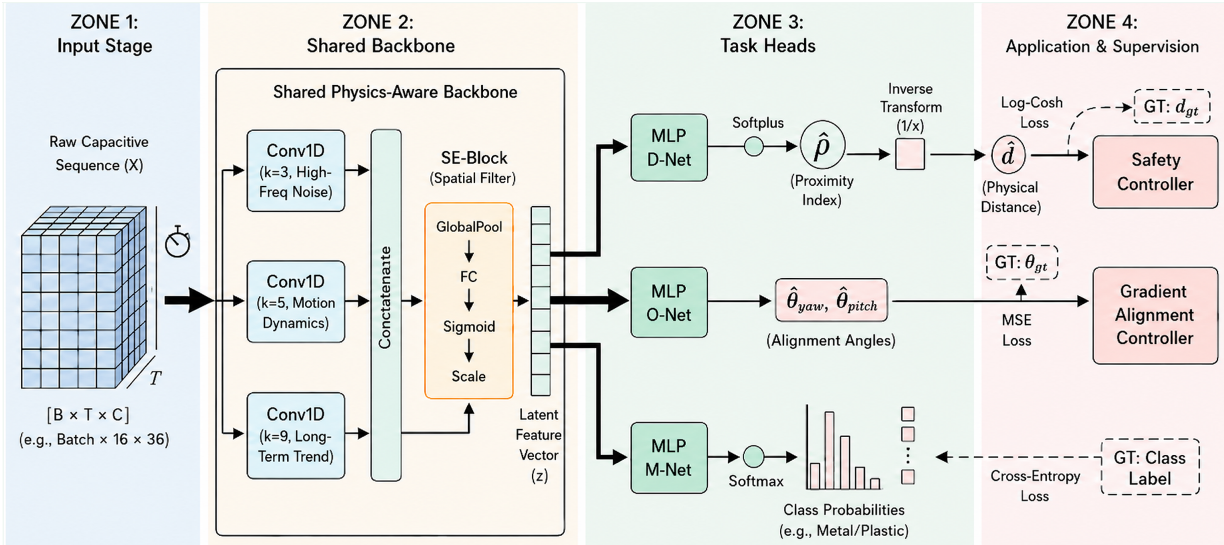

Fig. 4: Framework of CapacitiveServo-Net

The overall training objective is written as a weighted combination of the active task losses:

$$\mathcal{L}_{total} = \lambda_{prox}\mathcal{L}_{prox} + \lambda_{align}\mathcal{L}_{align} + \lambda_{cls}\mathcal{L}_{cls}. \quad (7)$$

The weights λ_{prox}, λ_{align}, and λ_{cls} are set according to the available annotations and the target deployment scenario. For example, the safety-skin experiments activate the proximity and alignment losses, whereas the gripper experiments activate the alignment and material-classification losses. This formulation allows the same backbone and decoder interface to be reused across sensor configurations while avoiding unnecessary loss terms for tasks without labels.

B. Angle Servoing for Pose Alignment and Distance Servoing for Safety Control

The outputs of CapacitiveServo-Net are used to generate closed-loop motion commands before physical contact occurs. Rather than relying on discrete proximity thresholds, we formulate a continuous control interface that combines pre-touch pose alignment with proximity-dependent velocity modulation. The estimated orientation $\hat{\boldsymbol{\theta}}$ is used for angular correction, while the proximity index $\hat{\rho}$ modulates translational motion and activates a repulsive safety response when the robot enters a predefined safety region.

1) Pre-touch Pose Alignment: Before contact, the robot aligns the sensor-equipped end-effector with the target surface using the predicted spatial misalignment $\hat{\boldsymbol{\theta}}$ between the sensor and the target. Here, $\hat{\boldsymbol{\theta}}$ is defined in the sensor frame and represents the relative angular offset from the aligned configuration. As shown in Fig. 5(a), the angular velocity command is given by

$$\boldsymbol{\omega}_{cmd} = -\mathbf{K}_{\omega}\hat{\boldsymbol{\theta}}, \quad (8)$$

where $\mathbf{K}_{\omega}$ is a positive-definite gain matrix. This servoing law reduces the relative angular misalignment between the sensor plane and the target surface, thereby improving the effective sensing geometry for subsequent proximity regulation.

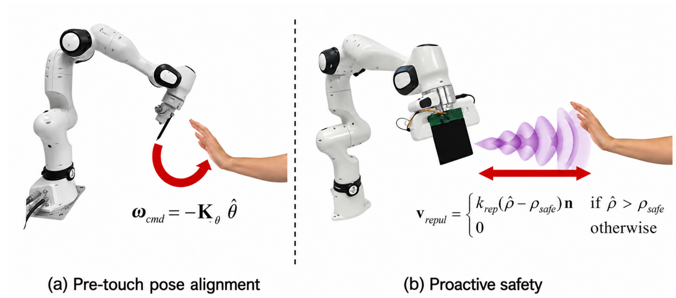

Fig. 5: Application of the *CapacitiveServo-Net* framework on a $10\,\text{cm} \times 10\,\text{cm}$ sensor-equipped robotic end-effector for pre-touch alignment and safety control.

2) Proximity-Based Safety Control: The predicted proximity index $\hat{\rho}$ is used to regulate the translational command through two complementary terms, as illustrated in Fig. 5(b). The first term attenuates the nominal velocity as the robot approaches an obstacle, while the second term introduces a repulsive response once the estimated proximity exceeds a safety threshold.

a) Proximity-dependent damping: Given the nominal velocity command $\mathbf{v}_{nom}$ from the high-level planner, we define a proximity-dependent scaling factor $\alpha(\hat{\rho}) \in [0, 1]$:

$$\mathbf{v}_{damp} = \alpha(\hat{\rho})\mathbf{v}_{nom}, \qquad \alpha(\hat{\rho}) = 1 - \hat{\rho}. \quad (9)$$

As the estimated proximity increases, this term attenuates the nominal velocity command, reducing the approach speed near the sensed object.

b) Repulsive safety response.: To maintain a safety margin, we define a threshold $\rho_{safe} = \mathcal{T}(d_{safe})$ corresponding to the desired safety distance d_{safe}. When $\hat{\rho} > \rho_{safe}$, a repulsive velocity is generated along the outward surface normal $\mathbf{n}$:

$$\mathbf{v}_{repel} = \begin{cases} k_{rep}(\hat{\rho} - \rho_{safe})\mathbf{n}, & \text{if } \hat{\rho} > \rho_{safe}, \\ \mathbf{0}, & \text{otherwise}, \end{cases} \quad (10)$$

where k_{rep} is a positive gain. The final translational command is

$$\mathbf{v}_{cmd} = \mathbf{v}_{damp} + \mathbf{v}_{repel}. \quad (11)$$

Together, these two terms slow the robot as it approaches nearby objects and actively bias the motion away from regions that violate the safety margin.

C. Material Recognition Application

To evaluate the versatility of the proposed framework beyond wide-area safety monitoring, we deploy *CapacitiveServo-Net* on a gripper-integrated ECT sensor, as shown in Fig. 6. This setting uses the same perception pipeline described in Sec. IV-A3, but adapts the input topology and task heads to support fine manipulation. Specifically, the gripper module uses a 2×4 electrode array with an active sensing area of $2\,\mathrm{cm} \times 9\,\mathrm{cm}$, resulting in $C = 28$ mutual-capacitance channels instead of the $C = 36$ channels used by the 3×3 Safety Skin.

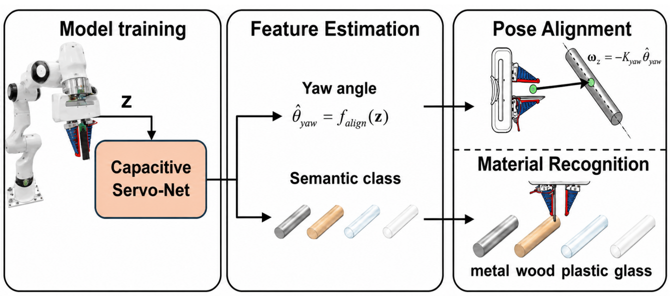

Fig. 6: Application of the *CapacitiveServo-Net* framework on a 2 cm$\times$9 cm sensor-equipped robotic gripper, demonstrating topology adaptation for fine manipulation.

The shared spatio-temporal backbone is retained across both sensor configurations, while the input layer and task-specific heads are instantiated according to the deployment scenario. For the gripper configuration, we activate the Alignment Head and Property Head in parallel. The Alignment Head estimates the yaw misalignment $\hat{\theta}_{yaw}$ between the gripper and the target object:

$$\hat{\theta}_{yaw} = f_{align}(\mathbf{z}), \quad (12)$$

which is used to generate an angular correction command,

$$\omega_z = -k_{yaw}\hat{\theta}_{yaw}. \quad (13)$$

This pre-touch servoing behavior aligns the gripper with the object's principal axis before contact, improving the expected contact geometry for grasping.

In parallel, the Property Head predicts the material category from the temporal capacitive response encoded in the latent feature $\mathbf{z}$. Different materials induce distinct mutual-capacitance perturbation patterns as the gripper approaches the object, allowing the network to estimate a categorical distribution $\mathbf{p} \in \mathbb{R}^K$ over K material classes. The two heads operate in a single forward pass, enabling the robot to estimate object orientation and material properties concurrently during the approach phase. These predictions can then be used for material-aware grasp refinement before physical contact.

V. Evaluation

A. Implementation and Data Collection

We evaluate the proposed framework on a robotic testbed covering both safety-oriented proximity control and gripper-based pre-touch manipulation. The platform consists of a 7-DOF Franka Emika Panda manipulator equipped with the custom ECT sensing modules described in Sec. III. This section describes the hardware integration, data collection protocol, baseline methods, and real-time inference setup.

For comparison, we implement Support Vector Regression (SVR) and K-Nearest Neighbors (KNN, $k = 5$) as classical regression baselines. All models are trained and evaluated using the same train/validation/test splits and the same synchronized capacitance–pose data. For the baseline methods, statistical features are extracted from the temporal capacitance window to provide a fair comparison with the proposed temporal model. Additional offline evaluation results are provided in the Appendix, and the real-time robot demonstrations are included in the accompanying video.

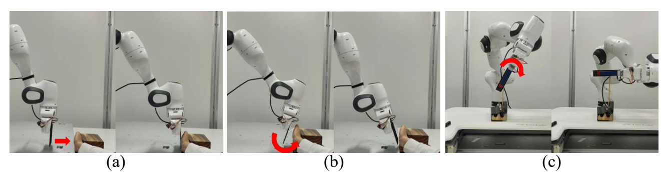

Fig. 7: Data collection process for proximity estimation and pre-touch manipulation tasks.

B. Evaluation of Distance Servoing for Safety Control

1) Offline Proximity Estimation: For the proximity estimation task, we execute a series of approach–retract trajectories against different target surfaces, as illustrated in Fig. 7(a). During each trajectory, the system records 36-channel mutual-capacitance measurements synchronized with ground-truth robot poses at a sampling frequency of 50 Hz. The collected dataset contains 984 interaction sequences, which are partitioned into training, validation, and test sets with an approximate ratio of 70:15:15.

To compare against non-temporal baselines, SVR and KNN are trained on hand-crafted statistical features, including the mean and standard deviation of each capacitance channel over the same 30-frame temporal window used by *CapacitiveServo-Net*. This setup evaluates whether end-to-end temporal feature learning improves proximity estimation compared with conventional feature-based regression. Additional quantitative results are reported in the Appendix.

2) *Real-Time Proximity Control:* We further evaluate the proposed model in a real-time closed-loop safety-control task. The manipulator approaches the target surface with a nominal velocity of $v_{nom} = 40\,\mathrm{mm/s}$, providing a dynamic setting for latency-sensitive proximity estimation and velocity modulation. During each trial, the models process the live sensor stream and generate distance predictions used for comparative control analysis.

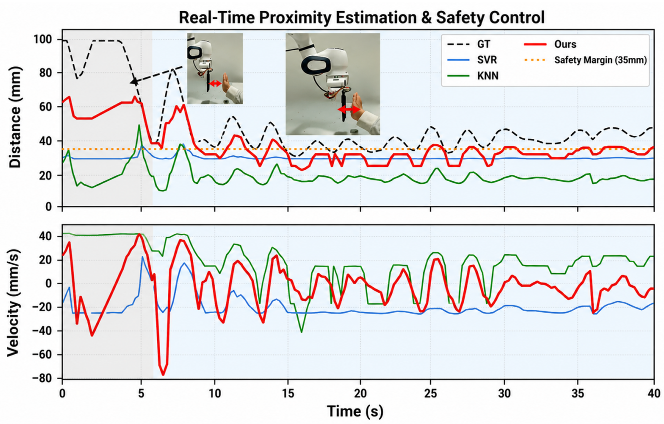

Fig. 8: Comparison of real-time proximity estimation and velocity modulation during closed-loop safety control.

As shown in Fig. 8, the safety margin is explicitly set to $35\,\mathrm{mm}$, corresponding to $\rho = 0.125$, and is indicated by the orange dotted line. This threshold provides a clear reference for evaluating how the controller modulates motion near the safety boundary. The predicted distance from *CapacitiveServo-Net* follows the overall trend of the ground-truth distance, with a small offset observed during close-range interaction. This offset reflects the proximity-based velocity modulation, which prioritizes gradual deceleration near the safety margin rather than instantaneous distance tracking. As the estimated proximity increases, the commanded approach velocity is attenuated, reducing the motion speed when the end-effector approaches the minimum distance margin.

The proposed *CapacitiveServo-Net* supports real-time inference within a 100 Hz closed-loop control pipeline. Compared with KNN, which tends to underestimate the distance and produces aggressive velocity fluctuations, and SVR, which exhibits a more delayed response, our model provides fixed-cost forward inference after training and generates responsive velocity commands near the safety boundary. The velocity profiles in Fig. 8 show that the proposed method can regulate the robot motion around the predefined safety margin, which is important for safety-oriented pHRI.

C. *Evaluation of Angle Servoing for Pose Alignment*

1) *Offline Pose Estimation:* To capture capacitive responses under varying orientations, we implement a Remote Center of Motion (RCM) controller that pivots the sensor about its surface center, as shown in Fig. 7(b). The pivot point is set at the sensor surface with a tool offset of $L = 0.09\,\mathrm{m}$, allowing the robot flange to rotate while keeping the sensor close to the target surface. The robot follows trajectories in Euler-angle space within $\pm 20^\circ$, covering diverse roll and pitch combinations. Data are recorded at 50 Hz, including 36-channel mutual-capacitance measurements and ground-truth robot poses. Raw signals are windowed with a sequence length of $N = 10$ to capture temporal dynamics. For the baselines, feature vectors are constructed from statistical moments, including mean and standard deviation, and temporal gradients Δ over the same window. The resulting pre-touch orientation dataset contains 5,537 samples and is split into training, validation, and test sets with an approximate ratio of 70:15:15.

2) *Real-Time Pose Alignment:* To evaluate dynamic feedback control, we conduct a real-time alignment experiment using only capacitive feedback. The objective is to keep the sensor-equipped end-effector parallel to a target surface, represented here by a human hand, from different initial angular deviations.

Experiment Setup and Implementation: A closed-loop velocity controller runs at 20 Hz. The predicted angular error θ_{err} is mapped to the angular velocity command $\omega_{cmd} = -K_p\theta_{err}$, with proportional gain $K_p = 0.5$ and exponential smoothing $\alpha = 0.1$ for safe interaction. To isolate single-axis behavior and improve visualization, we evaluate sequential pitch (Y-axis) and yaw (Z-axis) rotations. During each trial, the robot autonomously adjusts its end-effector orientation while the user keeps the hand approximately static.

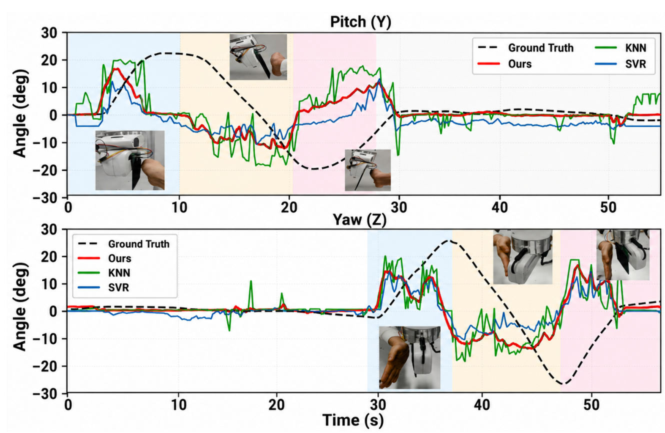

Fig. 9: Comparison of real-time pose alignment and control performance.

Visualization and Temporal Dynamics: Fig. 9 shows the real-time relationship between the human hand orientation and the robot's capacitive-feedback response. The model predicts the angular error from the sensor frame, whereas the ground-truth robot pose is recorded in the robot base frame. Therefore, the prediction and pose trajectories may exhibit opposite signs depending on the axis convention. In both pitch and yaw trials, changes in the predicted angular error appear before the corresponding robot motion, since the prediction serves as the control input while the physical response is delayed by robot dynamics and the smoothing filter.

Compared with KNN and SVR, *CapacitiveServo-Net* produces more temporally consistent angle estimates during closed-loop execution. KNN exhibits larger high-frequency fluctuations, while SVR tends to respond more conservatively to rapid orientation changes. In contrast, the proposed model better follows the hand-induced orientation changes and supports stable convergence toward the aligned configuration.

D. *Evaluation of Material Recognition Application*

1) Offline Angle and Material Estimation: We implement a contact-triggered acquisition strategy for collecting gripper data. As shown in Fig. 7(c), the gripper closes until a force threshold of $10\,\mathrm{N}$ or a collision event is detected, after which the system enters a stable holding state and records a 1-second temporal sequence of 50 frames at 50 Hz. The collected dataset contains 456 samples and is split into training, validation, and test sets with a 70:15:15 ratio. A stratified split is used to maintain balanced representation across the four material classes.

2) Real-Time Fine Grasping: To evaluate online applicability, we conduct real-time closed-loop angle servoing experiments using only capacitive feedback. The task is to rotate the gripper from a randomly initialized pre-grasp angle, approximately $10°$–$30°$, toward a target orientation of $90°$ before force closure. During this phase, the gripper remains non-contact with respect to the target object, and the controller relies solely on pre-touch capacitive measurements.

Experimental Setup and Implementation: We evaluate servoing performance across four material categories: wood, metal, plastic, and glass. In each trial, the robot approaches the object from a random initial angle without making contact. The controller uses the real-time angle prediction to compute a proportional velocity command and rotates the end-effector around a Remote Center of Motion (RCM) located at the intended grasp pivot. The control loop runs at 50 Hz, matching the temporal resolution of the training data.

Results and Discussion for Angle Servoing: The real-time pre-grasp angle servoing results are shown in Fig. 10. Across different materials and initial grasp angles, *CapacitiveServo-Net* tracks the target trajectory more consistently than the baseline methods using only non-contact capacitive feedback. For metal and plastic objects, the controller reaches the target $90°$ orientation before contact. For wood and glass, the controller still drives the grasp angle close to the target, reaching approximately $80.2°$ and $81.2°$, respectively. In contrast, KNN exhibits larger fluctuations and intermittent prediction drift, while SVR shows a more conservative response and often underestimates the required rotation. Despite material-dependent differences in capacitive response, the proposed model provides temporally consistent predictions for closed-loop RCM-based grasp servoing at 50 Hz.

These results suggest that *CapacitiveServo-Net* can support real-time capacitive servoing across different object materials and grasp configurations. Due to space constraints, real-time material classification results are provided in the Appendix.

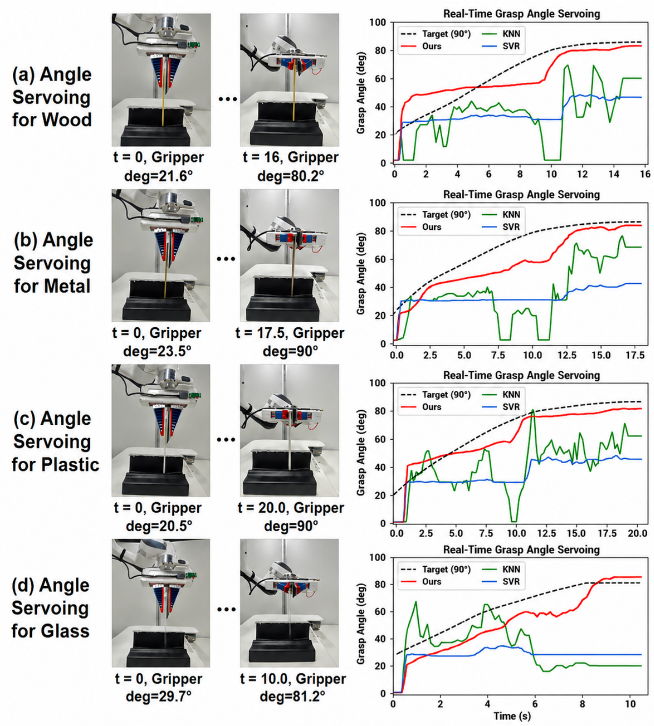

Fig. 10: Comparison of real-time fine grasping performance across wood, metal, plastic, and glass objects.

VI. Discussion and Conclusion

This work presents an ECT-based capacitive sensing framework for proactive robotic perception and control. Combined with *CapacitiveServo-Net*, open-boundary ECT sensors map mutual-capacitance perturbations directly to proximity, orientation, and material class. Experiments with planar and gripper-mounted sensors show that capacitive near-field perception bridges far-field vision and post-contact tactile feedback, enabling pre-contact motion adjustment for safety monitoring and pre-grasp manipulation.

Unlike optical proximity sensors, low-frequency capacitive fields are robust to lighting, reflectance, and line-of-sight limitations. By learning latent dielectric features directly from mutual-capacitance measurements, the method avoids tomographic reconstruction while supporting real-time servoing for safety regulation, pose alignment, and material-aware grasp refinement. Remaining challenges include ambiguity in multi-object or multi-material scenes and decoupling permittivity from contact force after interaction. Future work will explore domain randomization, online adaptation, hybrid force–proximity sensing, and integration with vision-language-action models under occlusion and visual degradation.

VII. Acknowledgment

This work was jointly supported by the National Natural Science Fund (62273054, 62120106005) and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under grant JYB2025XDXM109.

References

[1] Iris Andrussow, Huanbo Sun, Katherine J Kuchenbecker, and Georg Martius. Minsight: A fingertip-sized vision-based tactile sensor for robotic manipulation. *Advanced Intelligent Systems*, 5(8):2300042, 2023.

[2] Marco Costanzo, Giuseppe De Maria, Gaetano Lettera, and Ciro Natale. A multimodal approach to human safety in collaborative robotic workcells. *IEEE Transactions on Automation Science and Engineering*, 19(2):1202–1216, 2022. doi: 10.1109/TASE.2020.3043286.

[3] Ziqiang Cui, Yu Sun, Lifeng Zhang, and Huaxiang Wang. Planar electrical capacitance tomography dynamic imaging for non-destructive test. *IEEE Transactions on Instrumentation and Measurement*, 71:1–9, 2022.

[4] Serkan Ergun, Yitao Ding, Hosam Alagi, Christian Schöffmann, Barnaba Ubezio, Gergely Sóti, Michael Rathmair, Stephan Mühlbacher-Karrer, Ulrike Thomas, Björn Hein, Michael W. Hofbaur, and Hubert Zangl. A unified perception benchmark for capacitive proximity sensing towards safe human-robot collaboration (hrc). *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3634–3640, 2021. URL https://api.semanticscholar.org/CorpusID:239041217.

[5] Zackory Erickson, Henry M Clever, Vamsee Gangaram, Eliot Xing, Greg Turk, C Karen Liu, and Charles C Kemp. Characterizing multidimensional capacitive servoing for physical human–robot interaction. *IEEE Transactions on Robotics*, 39(1):357–372, 2022.

[6] Bernard Espiau, François Chaumette, and Patrick Rives. A new approach to visual servoing in robotics. In *Workshop on Geometric Reasoning for Perception and Action*, pages 106–136. Springer, 1991.

[7] Diogo Fonseca, Mohammad Safeea, and Pedro Neto. A flexible piezoresistive/self-capacitive hybrid force and proximity sensor to interface collaborative robots. *IEEE Transactions on Industrial Informatics*, 19(3):2485–2495, 2022.

[8] Gabriel J Garcia, Juan A Corrales, Jorge Pomares, and Fernando Torres. Survey of visual and force/tactile control of robots for physical interaction in spain. *Sensors*, 9(12):9689–9733, 2009.

[9] Boris G Gorshkov, Kivilcim Yüksel, Andrei A Fotiadi, Marc Wuilpart, Dmitry A Korobko, Andrey A Zhirnov, Konstantin V Stepanov, Artem T Turov, Yuri A Konstantinov, and Ivan A Lobach. Scientific applications of distributed acoustic sensing: State-of-the-art review and perspective. *Sensors*, 22(3):1033, 2022.

[10] Kaijen Hsiao, Paul Nangeroni, Manfred Huber, Ashutosh Saxena, and Andrew Y Ng. Reactive grasping using optical proximity sensors. In *2009 IEEE International Conference on Robotics and Automation*, pages 2098–2105. IEEE, 2009.

[11] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[12] Jianhao Jiao, Feiyi Chen, Hexiang Wei, Jin Wu, and Ming Liu. Lce-calib: Automatic lidar-frame/event camera extrinsic calibration with a globally optimal solution. *IEEE/ASME Transactions on Mechatronics*, 28(5):2988–2999, 2023.

[13] Xiaofei Liu and Wuqiang Yang. Measurement of liquid-level in horizontal pipe using 4-electrode ect sensor on robot hand. In *2023 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 1–5. IEEE, 2023.

[14] Quan Khanh Luu, Dinh Quang Nguyen, Nhan Huu Nguyen, Nam Phuong Dam, and Van Anh Ho. Vision-based proximity and tactile sensing for robot arms: Design, perception, and control. *IEEE Transactions on Robotics*, 41:5000–5019, 2025. doi: 10.1109/TRO.2025.3593087.

[15] Ke Ma, Xiaodong Chen, Jie Zhang, Zekai Xie, Jianing Wu, and Jinxiu Zhang. Inspired by physical intelligence of an elephant trunk: Biomimetic soft robot with pre-programmable localized stiffness. *IEEE robotics and automation letters*, 8(5):2898–2905, 2023.

[16] Wenye Ouyang, Liang He, Alessandro Albini, and Perla Maiolino. A modular soft robotic arm with embedded tactile sensors for proprioception. In *2022 IEEE 5th International Conference on Soft Robotics (RoboSoft)*, pages 919–924. IEEE, 2022.

[17] Samarathunga SMBPB, Marcello Valori, Giovanni Legnani, and Irene Fassi. Assessing safety in physical human–robot interaction in industrial settings: A systematic review of contact modelling and impact measuring methods. *Robotics*, 14(3):27, 2025.

[18] Joshua R Smith, Eric Garcia, Ryan Wistort, and Ganesh Krishnamoorthy. Electric field imaging pretouch for robotic graspers. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 676–683. IEEE, 2007.

[19] Xing Xu, Jie Li, Yang Yang, and Fumin Shen. Toward effective intrusion detection using log-cosh conditional variational autoencoder. *IEEE Internet of Things Journal*, 8(8):6187–6196, 2021. doi: 10.1109/JIOT.2020.3034621.

[20] Hongsik Yim, Hyunchang Kang, Seungjae Moon, Yeeun Kim, Tien Dat Nguyen, and Hyouk Ryeol Choi. Multi-functional safety sensor coupling capacitive and inductive measurement for physical human–robot interaction. *Sensors and Actuators A: Physical*, 354:114285, 2023.

[21] Maomao Zhang, Lu Ma, and MJFM Soleimani. Dual modality ect–mit multi-phase flow imaging. *Flow Measurement and Instrumentation*, 46:240–254, 2015.