TOUCH2PREDICT: LEARNING PHYSICAL PROPERTIES FROM A SINGLE TACTILE INTERACTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Estimating physical properties of deformable objects is essential for predictive robotic manipulation, but remains difficult in contact-rich interactions where visual deformation is small, localized, and often occluded. Tactile sensing provides direct measurements of mechanical response, yet these observations are highly sensitive to uncertain robot–object contact configuration, coupling pose and material estimation. We introduce Touch2Predict, a structured tactile inverse-modeling framework that separates these two factors according to their physical roles. Contact-establishment cues guide a low-dimensional derivative-free search over contact-relevant pose variables, while the full spatiotemporal pressure sequence drives adjoint-based material refinement through differentiable contact simulation. Alternating the two stages progressively recovers the interaction-specific contact state and object-level physical properties. Across eight synthetic scenes with translational, rotational, and coupled pose uncertainty, TOUCH2PREDICT recovers Young's modulus with an average per-scene median relative error of 6.05%. Real-robot experiments further demonstrate its practical feasibility, with the modulus identified for a test object within 12.4% of an independent reference measurement. Together with qualitative predictions under unseen interactions, these results highlight the potential of tactile inverse physics to recover transferable physical models for understanding and predicting physical interactions.

1 INTRODUCTION

Reliable robotic manipulation of deformable objects in contact-rich scenarios critically hinges on physical properties, including compliance, stiffness, damping, and friction, which shape contact-induced deformation and force transmission.

Vision-based inverse-physics methods infer physical properties from observed object dynamics (Li et al., 2023; Cai et al., 2024; Zhong et al., 2024; Chen et al., 2025a; Rho et al., 2026; Wang et al., 2026). Their evaluations typically consider large, visually salient deformations under free fall, passive collision, or externally applied forces. Such observational conditions rarely hold in contact-rich robotic manipulation, where deformations can be small, localized near the contact interface, and partially occluded by the robot. Moreover, the observed deformation is jointly determined by the object's physical properties and the robot–object contact configuration, rendering the underlying physical properties difficult to disentangle from vision alone, as illustrated in Figure 1.

Alternatively, tactile sensing provides direct observations of the mechanical response elicited during robot-object interaction. Through a controlled tactile interaction, the robot induces the response, which is captured as a dense, spatially resolved pressure sequence across the contact interface. However, localized pressure responses remain highly sensitive to the contact configuration: minor errors in calibration, depth estimation, object placement, or registration can perturb contact onset, effective indentation, and the load-bearing region. Sensitivity to contact configuration thus creates two intertwined difficulties: **First**, contact state and physical properties become confounded in the inverse problem, since variations in contact configuration can be partially compensated by changes in physical properties and yield similar pressure observations. **Second**, pose perturbations can alter contact onset and the active contact set, making local gradients unreliable for pose refinement. Consequently, directly optimizing both against the same raw-pressure objective may reproduce the observed tactile response while recovering erroneous physical properties.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

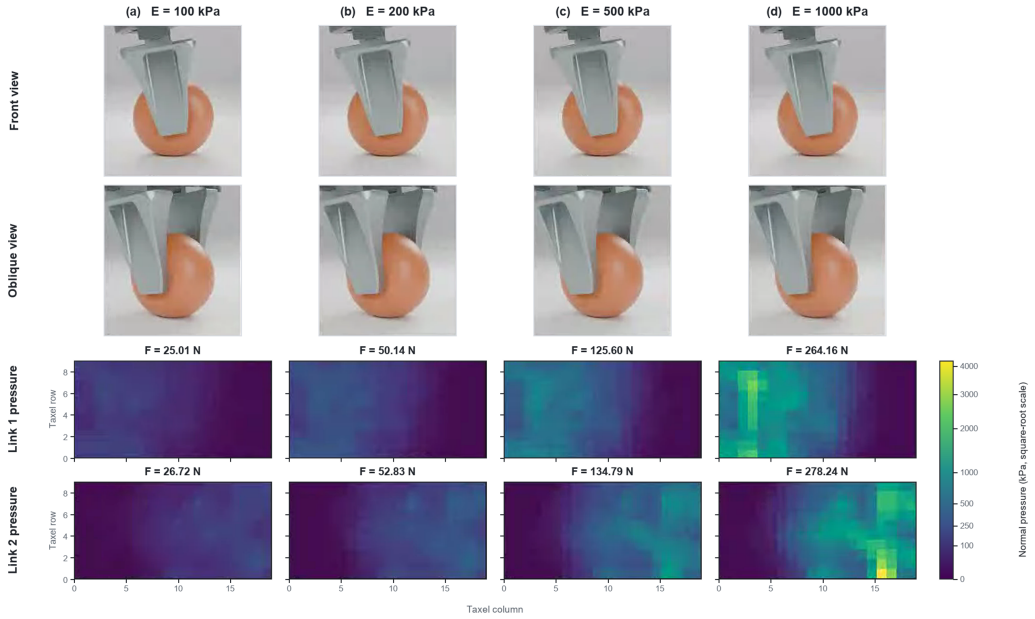


Figure 1: Similar visual deformation but stiffness-dependent tactile responses. For the same object geometry, robot motion, and initial conditions, increasing Young's modulus from 100 to 1000 kPa (a–d) yields stronger, more spatially concentrated contact pressure and an approximately tenfold increase in resultant contact force.

To address these difficulties, we introduce TOUCH2PREDICT, a cyclic optimization framework for tactile physical property estimation. Our key insight is to exploit the distinct roles of pose and physical properties: the interaction-specific pose determines where and when contact occurs, while object-level physical properties shape how load develops after contact. Rather than optimizing both against the same raw-pressure objective, TOUCH2PREDICT alternates between two optimization stages. The first stage refines contact-relevant pose variables through derivative-free search guided by contact-establishment timing and spatial contact structure. The second stage optimizes physical properties with adjoint differentiation, which efficiently computes gradients incorporating their influence throughout the simulated interaction. Repeating these two stages in a loop progressively refines both the interaction-specific contact state and the object-level physical model, thereby mitigating pose-physics compensation and enabling accurate physical property estimation from tactile sensing.

We evaluate TOUCH2PREDICT across eight synthetic scenes with translational, rotational, and coupled contact uncertainty. Under known contact geometry, a single probing sequence provides sufficient tactile information for material recovery in the tested settings. Across the evaluated scenes, TOUCH2PREDICT consistently achieves better material recovery than baselines. We further evaluate held-out interactions to test whether the recovered parameters capture transferable object-level physics rather than interaction-specific fits.

Our contributions are:

1. We formulate tactile physical-property identification under uncertain contact geometry and identify pose–physics compensation as a central ambiguity in physical recovery.
2. We introduce a structured alternating inference procedure that uses contact-establishment evidence for derivative-free contact-state search and the complete pressure response for adjoint-based material refinement.
3. We develop a GPU-accelerated C++/CUDA differentiable IPC framework that couples robot-driven deformable contact, dense tactile pressure simulation, and reverse-time adjoint optimization, and evaluate it across diverse contact uncertainties and held-out interactions

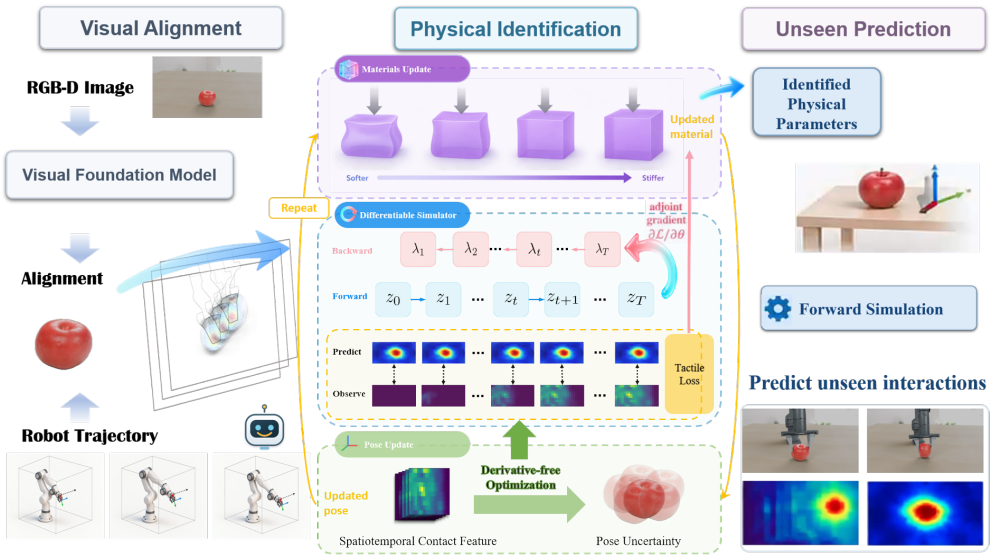


Figure 2: Overview of TOUCH2PREDICT. Starting from visual geometry and a coarse pose, the method alternates contact-structure-based pose refinement with adjoint-based material identification from a single tactile probing sequence. The recovered material parameters are then fixed for prediction under unseen interactions.

2 RELATED WORK

2.1 PHYSICAL PROPERTY IDENTIFICATION FROM DYNAMIC OBSERVATIONS

Recovering physical properties from dynamic observations is central to building predictive models of deformable-object manipulation. Vision-based inverse-physics methods infer material and dynamical parameters from observed motion and deformation using RGB images or videos, masks, optical flow, and 3D observations, often together with differentiable physical simulation (Li et al., 2023; Cai et al., 2024; Zhong et al., 2024; Chen et al., 2025a; Jiang et al., 2025; Yang et al., 2025; Rho et al., 2026; Wang et al., 2026). Complementary learning-based approaches learn predictive deformable-object dynamics or directly infer material properties from visual or RGB-D interaction data (Zhang et al., 2025; Huang et al., 2026a; Chen et al., 2026; Yin et al., 2026; Le et al., 2025; Huang et al., 2026b). While these methods exploit visible motion as physical evidence, contact-rich robotic manipulation poses additional challenges because informative deformation may be localized near the contact interface or occluded by the robot.

Tactile and proprioceptive observations provide more direct measurements of contact loads and mechanical response. Prior work has used tactile sequences for supervised material estimation, differentiable tactile simulation for sensor and physical calibration, visuo-tactile representations for deformable-object dynamics, and robot proprioception for physical-parameter identification (Burgess & Zhao, 2024; Si et al., 2024; Wi et al., 2022; Chen et al., 2025b). These approaches, however, generally assume that the interaction geometry is known or do not explicitly recover object-level constitutive parameters together with uncertain robot–object pose. Our setting instead identifies material parameters from a single dense tactile probing sequence while simultaneously refining the interaction-specific relative pose, addressing the resulting *pose–physics compensation* ambiguity.

2.2 DIFFERENTIABLE CONTACT DYNAMICS AND TACTILE SIMULATION

Existing differentiable deformable simulators mainly build on particle-based methods (Jiang et al., 2016) or mesh-based formulations. Differentiable MPM frameworks enable end-to-end gradients through large-deformation dynamics (Hu et al., 2019; 2020), while mesh-based methods explicitly represent deformable geometry and have also been developed for efficient differentiable simulation, such as DiffPD (Du et al., 2021). For contact-rich problems, IPC (Li et al., 2020) combines FEM

with barrier-based contact and continuous collision detection to robustly resolve non-penetrating contact. This explicit and robust treatment of surface contact is particularly well suited to tactile simulation, where observations are defined directly at the contact interface and depend on accurate surface geometry. GIPC (Huang et al., 2024a) and StiffGIPC (Huang et al., 2025) further accelerate IPC on GPUs for large-scale contact-rich simulation, while DiffIPC (Huang et al., 2024b) derives adjoint sensitivities through IPC dynamics for gradient-based inverse problems.

We therefore build our simulator on StiffGIPC, augment it with a lightweight tactile readout that projects contact forces onto sensor elements, and implement GPU-accelerated reverse-time adjoint differentiation following DiffIPC. This enables efficient optimization of constitutive parameters directly from temporal tactile observations.

2.3 Contact-Consistent Pose Refinement and Structured Optimization

Visual estimators can provide strong 6D pose initialization (Wen et al., 2024), but may lack the precision required for accurate contact modeling, especially under robot-induced occlusion. Refining these estimates through contact dynamics introduces further challenges: prior analyses identify gradient inaccuracies (Zhong et al., 2022) and rapidly varying local gradients in non-smooth contact dynamics (Suh et al., 2021).

Recent methods for rigid-object manipulation refine visual poses using tactile and proprioceptive constraints (Li et al., 2025), or estimate in-hand pose from force–torque observations through differentiable contact dynamics (Li et al., 2026). Deformable-object probing introduces an additional ambiguity, as pose and material can compensate for one another in the observed contact response. We therefore refine the visual initialization using low-dimensional, derivative-free search over only a task-constrained subset of contact-relevant translations and rotations, rather than re-estimating the full 6D pose. These contact-state updates alternate with adjoint-based material refinement to address pose–physics compensation.

3 Method

Overview. As illustrated in Figure 2, Touch2Predict identifies an object-level physical model from a single tactile pressure sequence by alternating pose and material refinement. Starting from visual geometry and a coarse pose, a shared robot–object simulator predicts tactile responses for each hypothesis. Each round first refines pose using the spatiotemporal structure of contact under the current material estimate.With pose fixed, a coarse material search over a short warm-up window initializes adjoint-based refinement against the complete pressure sequence. The updated material estimate then informs the next pose refinement. The identified material parameters are then kept fixed when predicting responses to unseen interactions.

3.1 Problem Formulation and Forward Model

We identify object-level material parameters from a single controlled robotic probing interaction. Given a visually reconstructed mesh $\mathcal{M}$, a coarse robot–object pose, a prescribed robot trajectory $\mathbf{u}_{0:T}$, and a dense tactile pressure sequence $\mathbf{Y}$, we estimate the constitutive parameters $\boldsymbol{\theta}$ while correcting a low-dimensional *contact-relevant pose* $\boldsymbol{\xi}$. The latter contains task-relevant in-plane translations, a prescribed-axis rotation, or their combination, rather than a full six-degree-of-freedom pose. It is interaction-specific, whereas $\boldsymbol{\theta}$ describes object-specific material properties within the validity of the constitutive model.

The predicted tactile sequence is

$$\widehat{\mathbf{Y}}(\boldsymbol{\xi}, \boldsymbol{\theta}) = \mathcal{H}_{\mathrm{tac}}[\mathcal{S}(\mathcal{M}, \mathbf{u}_{0:T}; \boldsymbol{\xi}, \boldsymbol{\theta})]\,, \tag{1}$$

where $\mathcal{S}$ is the robot–object simulator and $\mathcal{H}_{\mathrm{tac}}$ maps simulated normal contact forces to per-taxel pressures. We build on StiffGIPC (Huang et al., 2025), adding a lightweight tactile readout and GPU-accelerated reverse-time adjoints following DiffIPC (Huang et al., 2024b). The object is modeled with deformable FEM and the robot as an approximately rigid body driven by the prescribed trajectory. At each time step, the simulator updates its state through

$$\mathbf{z}_{t+1} = \arg\min_{\mathbf{z}} \frac{1}{2h^2}\|\mathbf{z} - \widetilde{\mathbf{z}}_t\|_{\mathbf{M}}^2 + \Psi_{\mathrm{el}}(\mathbf{z}; \boldsymbol{\theta}) + \Psi_{\mathrm{contact}}(\mathbf{z}; \mathbf{u}_t, \boldsymbol{\xi}), \tag{2}$$

where h is the time step, $\mathbf{M}$ is the mass matrix, $\widetilde{\mathbf{z}}_t$ is the inertial prediction, Ψ_{el} denotes material-dependent elastic energy, and Ψ_{contact} collects robot actuation, IPC barrier contact, and friction terms. All pose and material hypotheses use this shared forward model; the forward dynamics and tactile readout are detailed in Appendix A.1.

Direct pressure fitting can confuse geometric and material errors: changes in contact timing, indentation, and the loaded region can be partially compensated by an incorrect stiffness. To reduce this *pose–physics compensation*, Touch2Predict estimates contact pose from contact-establishment structure and material properties from the complete load-bearing pressure response.

3.2 Pose Refinement

Even with an accurate reconstructed object mesh, small errors in the initial object pose can substantially alter the simulated tactile response. Since all objects are stably supported on a tabletop, we optimize only the pose variables that affect the tactile interaction. Perturbations in these variables change contact onset, effective indentation, and the surface region that carries load, which can in turn bias subsequent material identification.

Directly minimizing the raw pressure residual is poorly suited to pose refinement because pressure magnitude is also strongly controlled by the material parameters. For example, a pose perturbation that induces excessive indentation may be partially offset by a softer material hypothesis, while insufficient indentation may be offset by a stiffer one. Consequently, a low pressure residual does not necessarily indicate accurate pose alignment when the material parameters are themselves uncertain.

We therefore refine the relevant pose variables as a separate optimization block before each material update. Rather than directly matching pressure amplitudes, we construct a structured pose objective from three complementary contact cues: contact evidence, contact-establishment timing, and the onset-aligned motion of the pressure center:

$$\begin{aligned} J_{\mathrm{evidence}} &= \left\langle (\delta_Z^{(f)})^2 \right\rangle_{\mathcal{F}}, \qquad J_{\mathrm{time}} = \left\langle (\delta_t^{(f)})^2 \right\rangle_{\mathcal{F}_c}, \\ J_{\mathrm{space}} &= \left\langle (\delta_c^{(f,j)})^2 \right\rangle_{\mathcal{P}}, \quad J_{\mathrm{pose}} = J_{\mathrm{evidence}} + J_{\mathrm{time}} + J_{\mathrm{space}}. \end{aligned} \tag{3}$$

Here, J_{evidence} measures discrepancies in contact evidence, J_{time} compares contact-establishment timing, and J_{space} compares the onset-aligned motion of the pressure center. The sets $\mathcal{F}$ and $\mathcal{F}_c$ index all tactile surfaces and surfaces with valid contact, respectively, while $\mathcal{P}$ indexes valid aligned surface–frame pairs. The normalized discrepancies δ_Z, δ_t, and δ_c measure the corresponding evidence, timing, and spatial errors, with D_f scaling center displacement by the taxel spacing. Complete definitions, including event detection, temporal alignment, and surface-balanced averaging, are provided in Appendix A.

Unlike the raw pressure objective, J_{pose} emphasizes the spatiotemporal structure of contact while reducing sensitivity to absolute pressure magnitude. It therefore retains the geometric cues needed to distinguish pose hypotheses under an imperfect material estimate. Candidates with inconsistent contact surfaces or invalid pressure-free alignment windows are additionally penalized.

Figure 3 illustrates this distinction over the same pose search region. Under a mismatched material hypothesis, the low-loss region of the raw pressure objective can shift away from the target pose, whereas the structured objective remains more closely aligned with the underlying contact configuration.

Because pose perturbations can change the active contact set, the resulting objective is generally nonsmooth and difficult to optimize reliably with gradients. We therefore use derivative-free CMA-ES to optimize J_{pose} within a bounded region around the initial pose estimate.

3.3 Adjoint-Based Material Identification

Conditioned on the current contact pose, we identify material parameters using the full pressure sequence:

$$J_{\mathrm{press}}(\boldsymbol{\xi}, \boldsymbol{\theta}) = \frac{\|\widehat{\mathbf{Y}}(\boldsymbol{\xi}, \boldsymbol{\theta}) - \mathbf{Y}\|_F^2}{2\max(\|\mathbf{Y}\|_F^2, \epsilon)}. \tag{4}$$

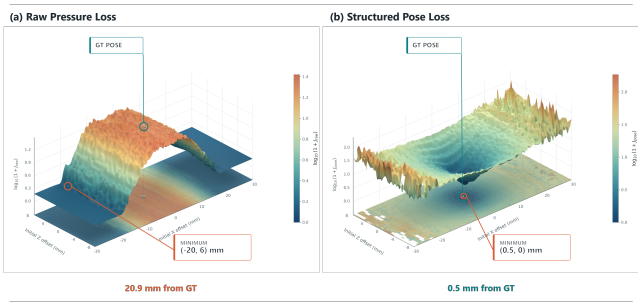


Figure 3: Pose-objective landscapes for a sphere with candidate $E = 1000$ kPa and target $E = 150$ kPa. (a) Raw pressure loss has a displaced low-loss basin and a minimum far from the ground-truth pose. (b) The structured pose loss combines contact evidence, establishment time, and onset-aligned pressure-center evolution, yielding a minimum near the ground-truth contact state.

Algorithm 1 Multi-step Touch2Predict Optimization

Inputs: mesh $\mathcal{M}$, robot motion $\mathbf{u}_{0:T}$, tactile sequence $\mathbf{Y}$
Hyperparameters: pose domain Ξ, material domain Θ, refinement rounds R
$\boldsymbol{\theta}^0 \leftarrow \text{Initialization}(\Theta)$
for $r = 0, \ldots, R-1$ **do**
  $\boldsymbol{\xi}^{r+1} \leftarrow \text{CMAES}_{\boldsymbol{\xi}\in\Xi}\, J_{\text{pose}}(\boldsymbol{\xi}; \boldsymbol{\theta}^r)$
  $\widetilde{\boldsymbol{\theta}}^{r+1} \leftarrow \text{CoarseSearch}_{\boldsymbol{\theta}\in\Theta}\, J^{\text{warm}}_{\text{press}}(\boldsymbol{\xi}^{r+1}, \boldsymbol{\theta})$
  $\boldsymbol{\theta}^{r+1} \leftarrow \text{LBFGS}_{\boldsymbol{\theta}}\left(J_{\text{press}}(\boldsymbol{\xi}^{r+1}, \boldsymbol{\theta}); \widetilde{\boldsymbol{\theta}}^{r+1}\right)$
end for
Return: $\boldsymbol{\xi}^R, \boldsymbol{\theta}^R$

The norm includes all tactile surfaces, time steps, and taxels, and $\epsilon > 0$ prevents division by zero. Unlike J_{pose}, this objective retains pressure magnitude to constrain the object's load-bearing response.

Let $\mathbf{R}(\mathbf{z}, \boldsymbol{\theta}; \boldsymbol{\xi}) = \mathbf{0}$ denote the discrete dynamics residual over the full trajectory, with $\mathbf{z}$ stacking the simulated states. For fixed $\boldsymbol{\xi}$, the adjoint system and material gradient are

$$\left(\frac{\partial \mathbf{R}}{\partial \mathbf{z}}\right)^{\top} \boldsymbol{\lambda} = \left(\frac{\partial J_{\text{press}}}{\partial \mathbf{z}}\right)^{\top}, \qquad \frac{\mathrm{d}J_{\text{press}}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\partial J_{\text{press}}}{\partial \boldsymbol{\theta}} - \boldsymbol{\lambda}^{\top} \frac{\partial \mathbf{R}}{\partial \boldsymbol{\theta}}. \tag{5}$$

A reverse adjoint pass therefore supplies material gradients without explicitly forming the full trajectory-to-parameter Jacobian. We primarily identify Young's modulus, optimizing $\eta = \log E$ to enforce positivity. With pose fixed, a coarse material search evaluates $J^{\text{warm}}_{\text{press}}$, the pressure objective restricted to a short, fixed warm-up window. The selected candidate initializes L-BFGS refinement of J_{press} over the complete pressure sequence, using the corresponding adjoint gradient.

Optimization Procedure. Algorithm 1 summarizes the complete optimization procedure.

4 Experiments

We evaluate Touch2Predict through three sets of experiments. First, under known contact geometry, we isolate material identification and compare inversion from visual and tactile observations. Second, we evaluate the complete inverse problem with uncertain contact state, including comparisons with alternative inference strategies and ablations of the proposed optimization procedure. We further examine whether the recovered physical parameters remain predictive under interactions that are not observed during identification. Finally, we validate the framework on a physical robot.

All simulated experiments use the same differentiable IPC-based robot–object simulation pipeline. Depending on the scene, the task-constrained contact state $\boldsymbol{\xi}$ contains translations, rotations, or their combination. Selected experiments additionally identify multiple physical parameters. For a compact comparison across heterogeneous scenes, the main text reports Young's-modulus recovery using

the median relative error. Detailed experimental settings and complete results are provided in the appendix.

4.1 Controlled Material Identification

We first isolate material identification from contact-state uncertainty by fixing the ground-truth contact geometry. We evaluate system identification from visual and tactile observations across several representative scenes, optimizing different physical parameters from the same prescribed initialization. The quantitative results are reported in Appendix Table 2, where each row lists the initial guess, the parameters obtained after optimization with each modality, and the corresponding ground truth.

Across all cases, tactile-based optimization recovers physical parameters close to the ground truth. In comparison, vision often struggles to resolve the small and localized changes induced by different physical parameters, leading to weaker and less stable optimization and, consequently, less accurate parameter recovery.

4.2 Complete Inverse-Physics Evaluation

We next evaluate the complete inverse problem, where the contact state is unknown and must be inferred together with the object's physical properties. Across scenes, the task-constrained contact state includes translation, rotation, or their combination. Selected settings additionally require simultaneous identification of multiple physical or contact parameters. Because these contact-state parameterizations have heterogeneous units and dimensions, we use Young's-modulus error as the primary metric in the main comparison and report detailed contact-state and multi-property results in the appendix.

Compared methods. We compare the complete method (**Full**) with three baselines. **Base** fixes the initial contact state and optimizes only material parameters. **Raw-Pose** retains Full's optimization procedure but replaces the structured contact objective with the full pressure residual. **Joint Raw CMA** jointly searches over contact-state and material parameters using CMA-ES with the same pressure objective. We also evaluate two ablations: **Full w/o warmup** removes the short-window coarse material warmup, while **Full w/o alt.** removes alternating contact-state and material updates. Optimization settings are detailed in Appendix C.2.3. An additional comparison with a higher-budget joint CMA-ES baseline is provided in Appendix C.2.7.

Material identification. Table 1 reports the median relative Young's-modulus error across the tested initialization conditions for each scene.

Table 1: Median material errors over completed runs across eight unknown-pose scenes. Material error is reported in percent; the final column is the unweighted mean of the eight displayed scene medians. Lower is better. Pale red and pale yellow mark the best and second-best result in each column, respectively.

	Material error (%)								
Method	Apple	Banana	Kong	Cup	Star_r	Banana_r	Torus	Sphere	Mean
Base	353.15	17.24	40.59	96.29	99.68	11.90	98.40	59.27	97.07
Raw-Pose	7.31	5.34	1.36	10.51	15.99	17.18	64.48	19.12	17.66
Joint Raw CMA	14.16	8.80	20.88	86.47	80.35	10.84	82.83	18.74	40.38
Full w/o warmup	2.84	6.42	1.50	9.27	10.02	28.76	54.41	1136.44	156.21
Full w/o alt.	4.03	1.91	1.10	10.44	14.87	13.64	34.04	**1.11**	10.14
Full	**1.87**	**0.97**	**1.00**	**7.80**	**4.92**	**3.79**	**26.87**	1.21	**6.05**

Optimization ablations. The ablations in Table 1 examine the two optimization components beyond the structured contact objective. Removing the warmup stage tests the sensitivity of local adjoint material refinement to its initialization, while removing alternation tests whether material updates improve subsequent contact-state estimation. Together, these comparisons isolate the contributions of the complete optimization procedure.

Prediction under unseen interactions. We keep the physical parameters identified from a single probing interaction fixed when simulating held-out interactions with different robot motions, contact locations, or contact configurations, using manually specified object poses. The qualitative comparisons in Figure 4 show that the recovered model more closely matches the ground-truth simulation than the initial model.

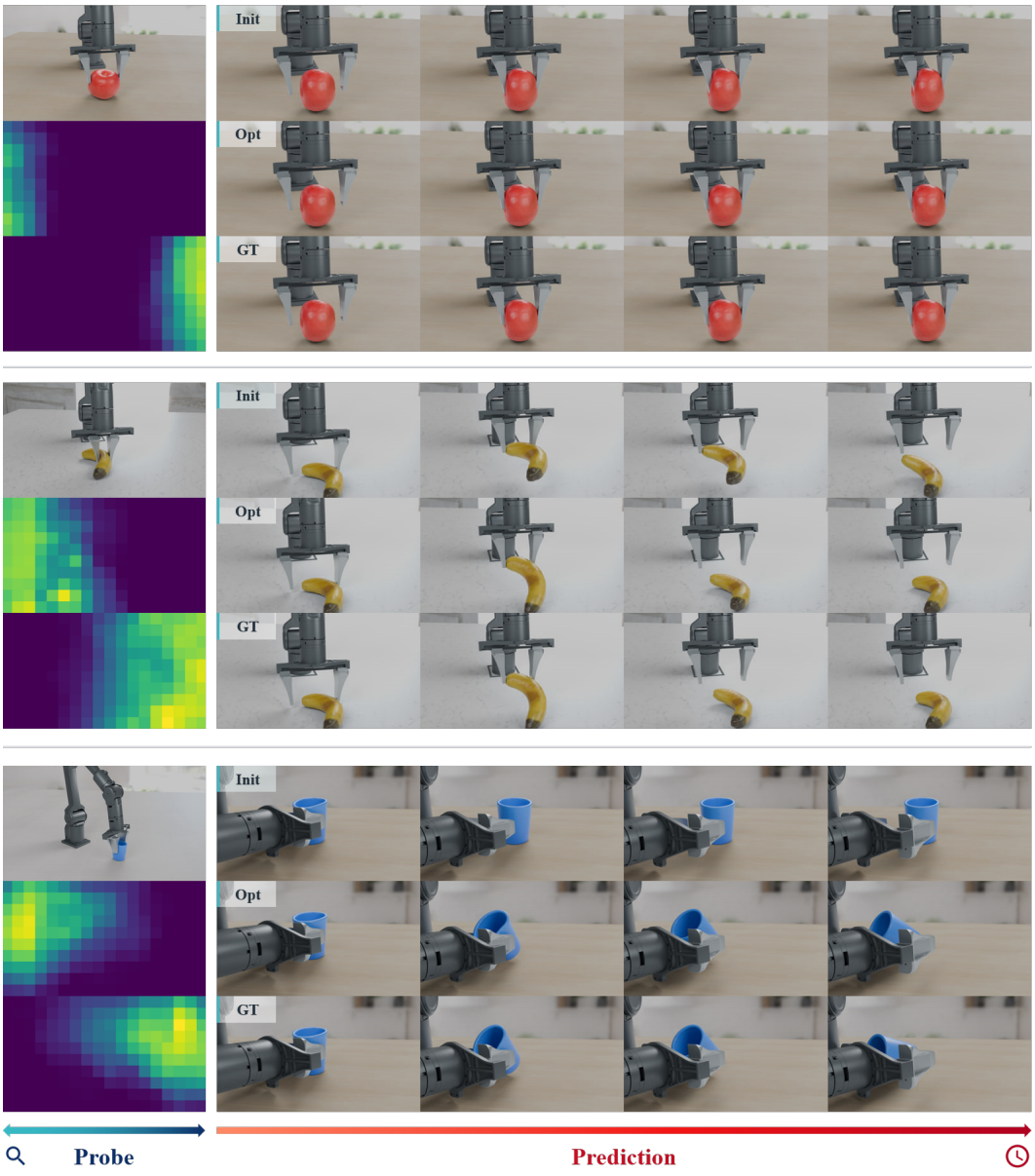


Figure 4: Qualitative prediction under unseen interactions. Physical parameters are identified from a single probing interaction and then frozen. Each row shows a held-out interaction with a robot motion, contact location, or contact configuration not used during optimization. We compare the ground-truth response with simulations using the initial physical model and the model recovered by Touch2Predict.

4.3 Real-Robot Experiments

We evaluate Touch2Predict on two real objects, a yellow egg and a blue ball; hardware and data-collection details are provided in Appendix C.3. For each object, we estimate Young's modulus from a single tactile probing interaction, resimulate that interaction, and predict the tactile response under a manually specified new object pose while keeping the estimated modulus fixed (Figure 5).

For the yellow egg, Touch2Predict estimates a Young's modulus of $0.1187\,\mathrm{MPa}$, compared with an independently measured third-party reference of $0.1056\,\mathrm{MPa}$, yielding a relative error of 12.4%. The simulated tactile responses show qualitative agreement with the measurements in both resimulation and the new-pose test. For the blue ball, we deliberately use an offset probing pose that produces localized edge contact during identification. Using the identified modulus, the new-pose prediction captures the transition to a broader contact region, although discrepancies in force magnitude and spatial distribution remain. Together, these results support the applicability of single-probing material identification to real objects and prediction beyond the contact configuration used for identification.

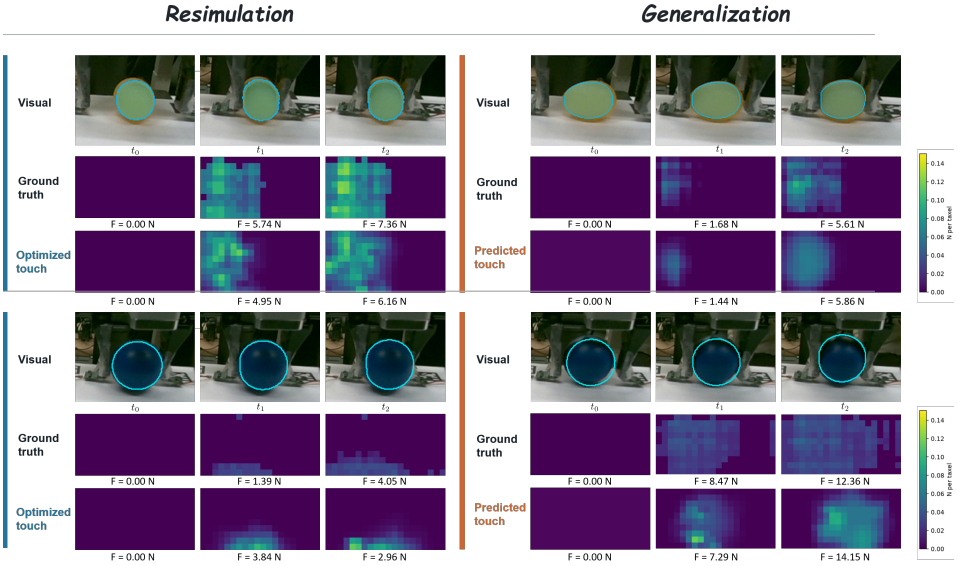


Figure 5: **Real-robot identification and prediction under new object poses.** Top: yellow egg. Bottom: blue ball. Left: resimulation of the probing interaction used for identification. Right: prediction under a new object pose with the identified Young's modulus held fixed. Within each block, columns show three representative time steps, and rows show visual observations, measured tactile responses, and resimulated or predicted tactile responses. For visualization, taxel pressures are converted to normal forces using the corresponding effective sensing areas. Colors indicate normal force per taxel (N). The blue-ball example exhibits similar visible contours but substantially different contact patterns between the two poses, illustrating the limited information that visual contours provide about the load-bearing contact configuration.

5 Conclusion

We presented Touch2Predict, a tactile inverse-physics framework for identifying object-level material properties from a single probing interaction under uncertain contact geometry. The method alternates structured, derivative-free pose refinement with adjoint-based material optimization through a shared differentiable simulator. Across eight synthetic scenes, the method reduces median material error relative to the three baselines. Through real-robot evaluation remains limited, qualitative held-out simulations and preliminary real-robot results suggest its potential for predicting object responses beyond the probing configuration.

AI USE STATEMENT

We used GPT to assist with drafting, translation, and language polishing of the manuscript, as well as brainstorming ideas for the proposed approach and interpreting selected experimental results. The authors critically evaluated the AI-assisted ideas and interpretations against relevant literature and experimental evidence, and reviewed and revised all AI-assisted text for accuracy and consistency with the work presented. All final methodological decisions and conclusions were made by the authors. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The potential ethical implications of this work depend on its application context. When applying the proposed method, practitioners should assess relevant risks, including safety, privacy, and fairness, and establish appropriate safeguards before deployment.

REPRODUCIBILITY STATEMENT

The proposed method is described in Section 3. The experimental settings and evaluation protocols are provided in Section 4, with additional implementation details in Appendix. These descriptions are intended to facilitate reproduction of the reported results. The code will be made publicly available upon acceptance.

REFERENCES

Michael Burgess and Jialiang Zhao. Learning object compliance via Young's modulus from single grasps with camera-based tactile sensors. arXiv:2406.15304, 2024. URL https://arxiv.org/abs/2406.15304.

Junhao Cai, Yuji Yang, Weihao Yuan, Yisheng He, Zilong Dong, Liefeng Bo, Hui Cheng, and Qifeng Chen. GIC: Gaussian-informed continuum for physical property identification and simulation. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/89379d5fc6eb34ff98488202fb52b9d0-Abstract-Conference.html.

Chuhao Chen, Zhiyang Dou, Chen Wang, Yiming Huang, Anjun Chen, Qiao Feng, Jiatao Gu, and Lingjie Liu. Vid2Sim: Generalizable, video-based reconstruction of appearance, geometry and physics for mesh-free simulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025a. URL https://czzzzh.github.io/Vid2Sim/.

Peter Yichen Chen, Chao Liu, Pingchuan Ma, John Eastman, Daniela Rus, Dylan Randle, Yuri Ivanov, and Wojciech Matusik. Learning object properties using robot proprioception via differentiable robot-object interaction. In *IEEE International Conference on Robotics and Automation*, 2025b. URL https://arxiv.org/abs/2410.03920.

Yunuo Chen, Yafei Hu, Lingfeng Sun, Tushar Kusnur, Laura Herlant, and Chenfanfu Jiang. Empm: Embodied mpm for modeling and simulation of deformable objects. *arXiv preprint arXiv:2601.17251*, 2026.

Tao Du, Kui Wu, Pingchuan Ma, Sebastien Wah, Andrew Spielberg, Daniela Rus, and Wojciech Matusik. DiffPD: Differentiable projective dynamics. *ACM Transactions on Graphics*, 41(2), 2021. doi: 10.1145/3490168.

Yuanming Hu, Jiancheng Liu, Andrew Spielberg, Joshua B. Tenenbaum, William T. Freeman, Jiajun Wu, Daniela Rus, and Wojciech Matusik. ChainQueen: A real-time differentiable physical simulator for soft robotics. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 6265–6271, 2019. doi: 10.1109/ICRA.2019.8794333.

Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. DiffTaichi: Differentiable programming for physical simulation. In *International Conference on Learning Representations (ICLR)*, 2020.

Kemeng Huang, Floyd M. Chitalu, Huancheng Lin, and Taku Komura. GIPC: Fast and stable gauss–newton optimization of IPC barrier energy. *ACM Transactions on Graphics*, 43(2), 2024a. doi: 10.1145/3643028. URL https://doi.org/10.1145/3643028.

Kemeng Huang, Xinyu Lu, Huancheng Lin, Taku Komura, and Minchen Li. StiffGIPC: Advancing GPU IPC for stiff affine-deformable simulation. *ACM Transactions on Graphics*, 44(3), 2025. doi: 10.1145/3735126.

Mu Huang, Hui Wang, Kerui Ren, Linning Xu, Yunsong Zhou, Mulin Yu, Bo Dai, and Jiangmiao Pang. SoMA: A real-to-sim neural simulator for robotic soft-body manipulation. arXiv:2602.02402, 2026a. URL https://arxiv.org/abs/2602.02402.

Qilin Huang, Quynh Anh Huynh, Long Le, Chen Wang, Chuhao Chen, Ryan Lucas, Eric Eaton, and Lingjie Liu. Unipixie: Unified and probabilistic 3d physics learning via flow matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026b.

Zizhou Huang et al. DiffIPC: Differentiable incremental potential contact. Project page and software release, 2024b.

Chenfanfu Jiang, Craig Schroeder, Joseph Teran, Alexey Stomakhin, and Andrew Selle. The material point method for simulating continuum materials. In *ACM SIGGRAPH 2016 Courses*, pp. 24:1–24:52. Association for Computing Machinery, 2016. doi: 10.1145/2897826.2927348.

Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, and Yunzhu Li. PhysTwin: Physics-informed reconstruction and simulation of deformable objects from videos. arXiv:2503.17973, 2025. URL https://arxiv.org/abs/2503.17973.

Long Le, Ryan Lucas, Wang Chen, Chuhao Chen, Dinesh Jayaraman, Eric Eaton, and Lingjie Liu. Pixie: Fast and generalizable supervised learning of 3d physics from pixels. *arXiv preprint arXiv:2508.17437*, 2025.

Hongyu Li, James Akl, Srinath Sridhar, Tye Brady, and Taskin Padir. ViTa-Zero: Zero-shot visuotactile object 6D pose estimation. arXiv:2504.13179, 2025. URL https://arxiv.org/abs/2504.13179.

Linfeng Li, Gang Yang, Lin Shao, and David Hsu. Differentiable contact dynamics for stable object placement under geometric uncertainties. *IEEE Robotics and Automation Letters*, 2026. URL https://arxiv.org/abs/2409.17725.

Minchen Li, Zachary Ferguson, Teseo Schneider, Timothy Langlois, Denis Zorin, Daniele Panozzo, Chenfanfu Jiang, and Danny M. Kaufman. Incremental potential contact: Intersection- and inversion-free large deformation dynamics. *ACM Transactions on Graphics*, 39(4), 2020. URL https://ipc-sim.github.io/.

Xuan Li, Yi-Ling Qiao, Peter Yichen Chen, Krishna Murthy Jatavallabhula, Ming Lin, Chenfanfu Jiang, and Chuang Gan. PAC-NeRF: Physics augmented continuum neural radiance fields for geometry-agnostic system identification. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tVkrbkz42vc.

Daniel Rho, Jun Myeong Choi, Matthew Thornton, Biswadip Dey, and Roni Sengupta. Monophysics: Estimating geometry, appearance, and physical parameters from monocular videos. *arXiv preprint arXiv:2605.30320*, 2026.

Zilin Si, Gu Zhang, Qingwei Ben, Branden Romero, Zhou Xian, Chao Liu, and Chuang Gan. DIFFTACTILE: A physics-based differentiable tactile simulator for contact-rich robotic manipulation. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=eJHnSg783t.

H. J. Terry Suh, Tao Pang, and Russ Tedrake. Bundled gradients through contact via randomized smoothing. arXiv:2109.05143, 2021. URL https://arxiv.org/abs/2109.05143.

Dan Wang, Xinrui Cui, Serge Belongie, and Ravi Ramamoorthi. Physconvex: Physics-informed 3d dynamic convex radiance fields for reconstruction and simulation. *arXiv preprint arXiv:2602.18886*, 2026.

Bowen Wen, Wei Yang, Jan Kautz, and Stan Birchfield. FoundationPose: Unified 6D pose estimation and tracking of novel objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17868–17879, 2024. URL https://arxiv.org/abs/2312.08344.

Youngsun Wi, Andy Zeng, Pete Florence, and Nima Fazeli. VIRDO++: Real-world, visuo-tactile dynamics and perception of deformable objects. arXiv:2210.03701, 2022. URL https://arxiv.org/abs/2210.03701.

Xintong Yang, Ze Ji, and Yu-Kun Lai. Differentiable physics-based system identification for robotic manipulation of elastoplastic materials. *The International Journal of Robotics Research*, 44(13): 2126–2155, 2025. doi: 10.1177/02783649251334661. URL https://doi.org/10.1177/02783649251334661.

Haocheng Yin, Shuohan Tao, Yongsheng Chen, and Lu Gan. Physcore: Physics-corrected residual world models for material-aware deformable dynamics. *arXiv preprint arXiv:2607.20653*, 2026.

Kaifeng Zhang, Baoyu Li, Kris Hauser, and Yunzhu Li. Particle-grid neural dynamics for learning deformable object models from RGB-D videos. In *Robotics: Science and Systems*, 2025. URL https://arxiv.org/abs/2506.15680.

Licheng Zhong, Hong-Xing Yu, Jiajun Wu, and Yunzhu Li. Reconstruction and simulation of elastic objects with spring-mass 3D gaussians. In *European Conference on Computer Vision*, 2024. URL https://zlicheng.com/spring_gaus/.

Yaofeng Desmond Zhong, Jiequn Han, and Georgia Olympia Brikis. Differentiable physics simulations with contacts: Do they have correct gradients w.r.t. position, velocity and control? arXiv:2207.05060, 2022. URL https://arxiv.org/abs/2207.05060.

A Method Derivations

This section provides the detailed definitions and derivations underlying the forward dynamics, structured pose objective, and adjoint-based material identification described in the main text. We use the same notation as in Sec. 3.1 and Sec. 3.3.

A.1 IPC Forward Dynamics

In Eq. (2), $\mathbf{z}_t$ denotes the configuration variables optimized by the forward solver. We use $\mathbf{s}_t$ for the complete simulator state, containing $\mathbf{z}_t$ and any auxiliary variables needed to make the discrete dynamics Markovian. The robot trajectory $\mathbf{u}_{0:T}$ is prescribed.

For the transition from t to $t+1$, the incremental potential is

$$\Phi_t(\mathbf{z}) = \frac{1}{2h^2}\|\mathbf{z} - \widetilde{\mathbf{z}}_t\|_{\mathbf{M}}^2 + \Psi_{\mathrm{el}}(\mathbf{z};\boldsymbol{\theta}) + \Psi_{\mathrm{contact}}(\mathbf{z};\mathbf{u}_t,\boldsymbol{\xi}), \tag{6}$$

where $\widetilde{\mathbf{z}}_t$ is the inertial prediction, Ψ_{el} is the elastic energy, and Ψ_{contact} collects robot actuation, IPC barrier contact, and friction terms. Dependence on the preceding complete state $\mathbf{s}_t$ is implicit in the notation; this state is held fixed within the current minimization. The next configuration is

$$\mathbf{z}_{t+1} = \arg\min_{\mathbf{z}} \Phi_t(\mathbf{z}). \tag{7}$$

At a converged forward step, the configuration residual satisfies

$$\mathbf{r}_{t+1}^{\mathrm{conf}} \equiv \nabla_{\mathbf{z}_{t+1}} \Phi_t(\mathbf{z}_{t+1}) = \mathbf{0}. \tag{8}$$

Equivalently,

$$\mathbf{r}_{t+1}^{\mathrm{conf}} = \frac{1}{h^2}\mathbf{M}\left(\mathbf{z}_{t+1} - \widetilde{\mathbf{z}}_t\right) + \nabla_{\mathbf{z}}\Psi_{\mathrm{el}}(\mathbf{z}_{t+1};\boldsymbol{\theta}) + \nabla_{\mathbf{z}}\Psi_{\mathrm{contact}}(\mathbf{z}_{t+1};\mathbf{u}_t,\boldsymbol{\xi}) = \mathbf{0}. \tag{9}$$

Combining this stationarity condition with the auxiliary state-update equations gives the complete dynamics residual

$$\mathbf{r}_t\left(\mathbf{s}_t, \mathbf{s}_{t-1}; \boldsymbol{\theta}, \boldsymbol{\xi}\right) = \mathbf{0}, \qquad t = 1, \ldots, T. \tag{10}$$

Here, $\mathbf{r}_t^{\mathrm{conf}}$ is the configuration component of $\mathbf{r}_t$. The complete state includes all propagated variables required by the time integrator and contact model, so the residual depends only on two consecutive complete states.

The corresponding state Jacobians are

$$\mathbf{A}_t = \frac{\partial \mathbf{r}_t}{\partial \mathbf{s}_t}, \qquad \mathbf{B}_t = \frac{\partial \mathbf{r}_t}{\partial \mathbf{s}_{t-1}}. \tag{11}$$

These linearize the complete dynamics residual, including both configuration stationarity and auxiliary state updates, and are used in Sec. A.3.

Tactile pressure readout. Let Ψ_{IPC} denote the IPC barrier contact potential used for tactile readout, expressed in physical energy units, and let $\mathbf{x}_{t,i}$ denote a sensor-surface vertex position. Nodal contact forces are obtained from its negative gradient and projected onto taxels:

$$\begin{aligned} \mathbf{f}_{t,i}^{\mathrm{c}} &= -\frac{\partial \Psi_{\mathrm{IPC}}}{\partial \mathbf{x}_{t,i}}, \\ \mathbf{F}_{t,k}^{(f)} &= \sum_i w_{t,ki}^{(f)} \mathbf{f}_{t,i}^{\mathrm{c}}, \qquad \widehat{P}_{t,k}^{(f)} = \frac{\left[-(\mathbf{n}_{t,k}^{(f)})^\top \mathbf{F}_{t,k}^{(f)}\right]_+}{A_k^{(f)}}. \end{aligned} \tag{12}$$

Here, $w_{t,ki}^{(f)}$ denotes the geometric vertex-to-taxel projection weight, $\mathbf{n}_{t,k}^{(f)}$ is the outward sensor normal, and $A_k^{(f)}$ is the effective sensing area. The positive part retains only compressive normal loads. Material optimization uses the resulting taxel pressures rather than per-taxel forces.

A.2 Structured Pose Objective

We next provide the complete definitions underlying J_{pose} in Eq. (3). The objective uses contact-establishment structure rather than directly matching absolute pressure magnitude. It consists of three complementary signals: contact evidence, establishment timing, and onset-aligned spatial motion of the pressure center.

Contact activation and establishment. For tactile surface f and taxel k, let $P_{t,k}^{(f)}$ denote the measured pressure at time t. The binary contact support is

$$a_{t,k}^{(f)} = \mathbf{1}\left[P_{t,k}^{(f)} > \tau\right], \tag{13}$$

where τ is the contact threshold. The newly activated support and its corresponding activation count are

$$e_{t,k}^{(f)} = \left[a_{t,k}^{(f)} - a_{t-1,k}^{(f)}\right]_+, \qquad A_t^{(f)} = \sum_k e_{t,k}^{(f)}, \tag{14}$$

where $[x]_+ = \max(x, 0)$ and support before the observation window is treated as zero.

To obtain a soft representation of the first contact-establishment event, we convert $A_t^{(f)}$ into an event probability using

$$h_t^{(f)} = 1 - \exp\left(-\frac{A_t^{(f)}}{N_{\text{ref}}}\right), \tag{15}$$

where N_{ref} is a fixed reference activation count. We set $N_{\text{ref}} = 40$ based on a sweep over $\{5, 20, 40, 60, 80, 120\}$ in a single scene. This value was selected as an empirical compromise between the temporal concentration of the soft first-contact event weights and the observed smoothness of the pose-loss curve, without claiming optimality. We then define

$$p_t^{(f)} = h_t^{(f)} \prod_{s<t} \left(1 - h_s^{(f)}\right), \qquad Z^{(f)} = \sum_t p_t^{(f)}. \tag{16}$$

Here, $p_t^{(f)}$ represents the soft first-event weight at time t, while $Z^{(f)}$ summarizes the amount of valid contact-establishment evidence observed on surface f.

For $Z^{(f)} > 0$, the normalized temporal weights are

$$\rho_t^{(f)} = \frac{p_t^{(f)}}{Z^{(f)}}, \tag{17}$$

and the contact-establishment time is defined as

$$\bar{t}^{(f)} = \sum_t \rho_t^{(f)} t. \tag{18}$$

Pressure-center trajectory. Let $\mathbf{r}_k^{(f)}$ denote the fixed surface-local coordinate of taxel k. For frames with positive total pressure, we define the pressure-weighted contact center as

$$\mathbf{c}_t^{(f)} = \frac{\sum_k P_{t,k}^{(f)} \mathbf{r}_k^{(f)}}{\sum_k P_{t,k}^{(f)}}. \tag{19}$$

For each tactile surface, candidate and target pressure sequences are aligned at their respective first positive-pressure frames. We denote these frames by $\widehat{t}_0^{(f)}$ and $t_0^{(f)}$, respectively, where hatted quantities correspond to the simulated candidate and unhatted quantities to the observed target.

Normalized discrepancies. The contact-evidence discrepancy is

$$\delta_Z^{(f)} = \frac{\widehat{Z}^{(f)} - Z^{(f)}}{s_Z}, \tag{20}$$

and the contact-establishment timing discrepancy is

$$\delta_t^{(f)} = \frac{\widehat{\bar{t}}^{(f)} - \bar{t}^{(f)}}{s_t}. \tag{21}$$

Here, s_Z and s_t provide fixed normalizations for the corresponding quantities.

For the spatial term, let D_f denote the diagonal matrix containing the taxel spacings of surface f. The onset-aligned spatial discrepancy at relative frame j is

$$\delta_c^{(f,j)} = \left\| D_f^{-1} \left(\widehat{\mathbf{c}}^{(f)}_{\widehat{t}_0^{(f)}+j} - \mathbf{c}^{(f)}_{t_0^{(f)}+j} \right) \right\|_2 . \tag{22}$$

The taxel-spacing normalization makes this discrepancy comparable across tactile surfaces with different physical resolutions.

Structured pose loss. Let $\mathcal{F}$ denote all tactile surfaces and $\mathcal{F}_c$ the subset with valid target contact. Let $\mathcal{P}$ denote the valid onset-aligned surface–frame pairs (f, j) used for spatial comparison. We use surface-balanced averaging: frames are first averaged within each valid surface and the resulting surface terms are then weighted equally.

With these definitions, the three components in Eq. (3) are

$$\begin{aligned} J_{\text{evidence}} &= \left\langle \left(\delta_Z^{(f)} \right)^2 \right\rangle_{\mathcal{F}}, \\ J_{\text{time}} &= \left\langle \left(\delta_t^{(f)} \right)^2 \right\rangle_{\mathcal{F}_c}, \\ J_{\text{space}} &= \left\langle \left(\delta_c^{(f,j)} \right)^2 \right\rangle_{\mathcal{P}}, \end{aligned} \tag{23}$$

and

$$J_{\text{pose}} = J_{\text{evidence}} + J_{\text{time}} + J_{\text{space}}. \tag{24}$$

The objective therefore emphasizes the spatiotemporal organization of contact rather than its absolute force magnitude. Candidates that do not produce the required contact support or do not admit a valid aligned pressure-center comparison are treated as invalid candidates in the pose search rather than assigning artificial event times or centers.

A.3 Adjoint-Based Material Gradient

With the current pose $\boldsymbol{\xi}$ fixed, material identification minimizes the pressure objective in Eq. (4). The derivation below first assumes a fixed initial state, $\mathrm{d}\mathbf{s}_0/\mathrm{d}\boldsymbol{\theta} = \mathbf{0}$. Derivatives are evaluated locally along a converged trajectory where the residual and tactile readout are differentiable and $\mathbf{A}_t$ is nonsingular.

Tactile observation gradient. Let $\widehat{\mathbf{y}}_t$ and $\mathbf{y}_t$ denote vectorized predicted and measured pressures over all tactile surfaces and taxels. Using the complete state notation,

$$\widehat{\mathbf{y}}_t = \mathcal{H}_{\text{tac}}(\mathbf{s}_t), \tag{25}$$

with dependence on the prescribed trajectory and fixed pose suppressed. Define

$$c_{\mathbf{Y}} = \max\left(\|\mathbf{Y}\|_F^2, \epsilon \right). \tag{26}$$

The pressure objective is

$$J_{\text{press}} = \frac{1}{2c_{\mathbf{Y}}} \sum_{t=1}^{T} \|\widehat{\mathbf{y}}_t - \mathbf{y}_t\|_2^2, \tag{27}$$

and its state derivative is

$$\left(\frac{\partial J_{\text{press}}}{\partial \mathbf{s}_t} \right)^\top = \left(\frac{\partial \widehat{\mathbf{y}}_t}{\partial \mathbf{s}_t} \right)^\top \frac{\widehat{\mathbf{y}}_t - \mathbf{y}_t}{c_{\mathbf{Y}}}. \tag{28}$$

Material sensitivity of the dynamics. Define the material Jacobian of the complete residual as

$$\mathbf{C}_t = \frac{\partial \mathbf{r}_t}{\partial \boldsymbol{\theta}}. \tag{29}$$

If $\boldsymbol{\theta}$ enters the dynamics explicitly only through the elastic energy, its configuration block is

$$\mathbf{C}_t^{\mathrm{conf}} \equiv \frac{\partial \mathbf{r}_t^{\mathrm{conf}}}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} \left[\nabla_{\mathbf{z}_t} \Psi_{\mathrm{el}}(\mathbf{z}_t; \boldsymbol{\theta}) \right], \tag{30}$$

and the auxiliary residual blocks have zero explicit material derivatives. Under this assumption, contact and friction still influence the material gradient through the states and state Jacobians.

Direct differentiation of the objective gives

$$\frac{\mathrm{d}J_{\mathrm{press}}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\partial J_{\mathrm{press}}}{\partial \boldsymbol{\theta}} + \sum_{t=1}^{T} \frac{\partial J_{\mathrm{press}}}{\partial \mathbf{s}_t} \frac{\mathrm{d}\mathbf{s}_t}{\mathrm{d}\boldsymbol{\theta}}, \tag{31}$$

while differentiating $\mathbf{r}_t = \mathbf{0}$ gives

$$\mathbf{A}_t \frac{\mathrm{d}\mathbf{s}_t}{\mathrm{d}\boldsymbol{\theta}} + \mathbf{B}_t \frac{\mathrm{d}\mathbf{s}_{t-1}}{\mathrm{d}\boldsymbol{\theta}} + \mathbf{C}_t = \mathbf{0}. \tag{32}$$

We eliminate these state sensitivities using a discrete adjoint.

Reverse-time adjoint. Introduce one adjoint variable per complete dynamics residual and define

$$\mathcal{L} = J_{\mathrm{press}} - \sum_{t=1}^{T} \boldsymbol{\lambda}_t^\top \mathbf{r}_t. \tag{33}$$

Along the converged trajectory, $\mathcal{L} = J_{\mathrm{press}}$. Each $\mathbf{s}_t$ appears in $\mathbf{r}_t$ and, for $t < T$, $\mathbf{r}_{t+1}$, giving

$$\left(\frac{\partial \mathcal{L}}{\partial \mathbf{s}_t} \right)^\top = \left(\frac{\partial J_{\mathrm{press}}}{\partial \mathbf{s}_t} \right)^\top - \mathbf{A}_t^\top \boldsymbol{\lambda}_t - \mathbf{B}_{t+1}^\top \boldsymbol{\lambda}_{t+1}. \tag{34}$$

Setting these derivatives to zero yields

$$\boxed{\mathbf{A}_t^\top \boldsymbol{\lambda}_t = \left(\frac{\partial J_{\mathrm{press}}}{\partial \mathbf{s}_t} \right)^\top - \mathbf{B}_{t+1}^\top \boldsymbol{\lambda}_{t+1}} \tag{35}$$

for $t = T, \ldots, 1$, using the terminal convention

$$\boldsymbol{\lambda}_{T+1} = \mathbf{0}, \qquad \mathbf{B}_{T+1} = \mathbf{0}. \tag{36}$$

In particular,

$$\mathbf{A}_T^\top \boldsymbol{\lambda}_T = \left(\frac{\partial J_{\mathrm{press}}}{\partial \mathbf{s}_T} \right)^\top. \tag{37}$$

The recursion combines the current pressure sensitivity with the effect of the current state on subsequent simulation steps.

Material gradient. With the adjoints chosen as above and the initial state fixed, the implicit state-sensitivity terms cancel, yielding

$$\boxed{\frac{\mathrm{d}J_{\mathrm{press}}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\partial J_{\mathrm{press}}}{\partial \boldsymbol{\theta}} - \sum_{t=1}^{T} \boldsymbol{\lambda}_t^\top \mathbf{C}_t}. \tag{38}$$

For the state-only tactile readout in Eq. (25), $\partial J_{\mathrm{press}}/\partial \boldsymbol{\theta} = \mathbf{0}$. Under the elastic-only parameter dependence described above, let $\boldsymbol{\lambda}_t^{\mathrm{conf}}$ denote the adjoint block paired with $\mathbf{r}_t^{\mathrm{conf}}$. Then

$$\frac{\mathrm{d}J_{\mathrm{press}}}{\mathrm{d}\boldsymbol{\theta}} = -\sum_{t=1}^{T} \left(\boldsymbol{\lambda}_t^{\mathrm{conf}} \right)^\top \frac{\partial}{\partial \boldsymbol{\theta}} \left[\nabla_{\mathbf{z}_t} \Psi_{\mathrm{el}}(\mathbf{z}_t; \boldsymbol{\theta}) \right]. \tag{39}$$

If the initial state depends on $\boldsymbol{\theta}$, the boundary term $-\boldsymbol{\lambda}_1^\top \mathbf{B}_1 (\mathrm{d}\mathbf{s}_0/\mathrm{d}\boldsymbol{\theta})$ must additionally be included in Eqs. (38) and (39).

Our implementation follows the discrete-adjoint formulation of DiffIPC (Huang et al., 2024b), with reverse-time solves executed on the GPU within the StiffGIPC (Huang et al., 2025) pipeline.

B Optimization and Implementation Details

This section summarizes the numerical settings used by the pose and material optimization stages in Algorithm 1. Both stages use the same forward simulator, with stage-specific objectives and optimization procedures.

B.1 Pose-Objective Evaluation

For the contact-support definition in Eq. (13), we use

$$\tau = 0. \tag{40}$$

Support before the observation window is treated as zero.

For the normalized discrepancies in Eqs. (20) and (21), we use

$$s_Z = 1, \qquad s_t = 1 \text{ frame}. \tag{41}$$

Candidate and target pressure sequences are aligned independently on each tactile surface using their first positive-pressure frames before the spatial term is evaluated. Spatial discrepancies are normalized by the physical taxel spacing through D_f.

The spatial loss is surface-balanced: valid aligned frames are first averaged within each tactile surface, and the resulting per-surface terms are subsequently averaged across valid surfaces. This prevents a surface with a longer valid contact interval from dominating J_{space}.

Pose candidates that fail to establish the required contact support, or for which a valid pressure-center alignment cannot be constructed, are treated as invalid and penalized by the pose search. We do not fabricate event times or pressure centers for pressure-free candidates.

Because changes in the contact-relevant pose may create, remove, or reorder active contacts, J_{pose} is generally nonsmooth with respect to $\boldsymbol{\xi}$. We therefore optimize it using CMA-ES over the bounded pose domain Ξ around the initial visual estimate, as described in Algorithm 1.

B.2 Material Parameterization

The primary material parameter in the main experiments is Young's modulus E. We optimize it in logarithmic coordinates,

$$\eta = \log E, \qquad E = \exp(\eta), \tag{42}$$

which enforces positivity throughout optimization.

The adjoint derivative with respect to E is converted to the optimization coordinate using

$$\frac{\mathrm{d}J_{\text{press}}}{\mathrm{d}\eta} = \frac{\mathrm{d}J_{\text{press}}}{\mathrm{d}E}\frac{\mathrm{d}E}{\mathrm{d}\eta} = E\frac{\mathrm{d}J_{\text{press}}}{\mathrm{d}E}. \tag{43}$$

The same construction applies elementwise when multiple positive constitutive parameters are optimized.

The admissible Young's-modulus range used by the optimizer is

$$E \in [10\,\text{kPa}, 100\,\text{MPa}]. \tag{44}$$

B.3 Coarse Material Initialization

Before continuous adjoint-based refinement, we perform a lightweight coarse search in log space. The Young's-modulus candidates are

$$\{10,\ 31.6,\ 100,\ 316,\ 1000,\ 3162,\ 10000,\ 31623,\ 100000\}\ \text{kPa}. \tag{45}$$

Each candidate is evaluated using the shared forward simulator and $J_{\text{press}}^{\text{warm}}$ over the same fixed warm-up window. The window is determined from the target sequence as specified in Sec. C.2.1, independently of candidate contact onset. Among numerically valid candidates, the one with the

lowest $J_{\text{press}}^{\text{warm}}$ initializes the subsequent L-BFGS refinement of J_{press} over the complete observation window.

For later material updates in an alternating trajectory, the material estimate inherited from the preceding refinement stage is additionally included among the coarse candidates. Candidates are deduplicated in log space before evaluation.

B.4 Adjoint-Based Material Refinement

Starting from the selected coarse-search candidate, we optimize J_{press} over the complete tactile sequence using L-BFGS. At every L-BFGS iterate, the forward simulator evaluates the pressure objective and the reverse-time adjoint in Eq. (35) provides the corresponding material gradient.

Native C++ L-BFGS/Armijo proposals are projected into the admissible material interval in Eq. (44) and replayed using the complete forward simulator. A candidate update is retained only when the forward result is finite, numerically valid, and strictly decreases the configured pressure objective.

The soft and hard material initializations used throughout the main experiments are fixed absolute values,

$$E_{\text{soft}} = 100\,\text{kPa}, \qquad E_{\text{hard}} = 10\,\text{MPa}. \tag{46}$$

These initializations are fixed independently of the ground-truth material parameter.

B.5 Alternating Pose and Material Optimization

The complete procedure follows Algorithm 1. At refinement round r, the current material estimate $\boldsymbol{\theta}^r$ is first held fixed while CMA-ES updates the contact-relevant pose:

$$\boldsymbol{\xi}^{r+1} = \text{CMAES}_{\boldsymbol{\xi}\in\Xi}\, J_{\text{pose}}(\boldsymbol{\xi}; \boldsymbol{\theta}^r). \tag{47}$$

The updated pose is then fixed while the material stage performs short-window coarse initialization followed by full-window adjoint-based L-BFGS refinement:

$$\widetilde{\boldsymbol{\theta}}^{r+1} = \text{CoarseSearch}_{\boldsymbol{\theta}\in\Theta}\, J_{\text{press}}^{\text{warm}}(\boldsymbol{\xi}^{r+1}, \boldsymbol{\theta}), \tag{48}$$

followed by

$$\boldsymbol{\theta}^{r+1} = \text{LBFGS}_{\boldsymbol{\theta}}\left(J_{\text{press}}(\boldsymbol{\xi}^{r+1}, \boldsymbol{\theta}); \widetilde{\boldsymbol{\theta}}^{r+1}\right). \tag{49}$$

The two optimization blocks intentionally use different information from the same tactile interaction. Pose refinement relies on contact evidence, establishment timing, and onset-aligned spatial support, while material refinement retains the full pressure magnitude over the load-bearing response. The alternating procedure therefore separates the geometric information used to establish contact configuration from the mechanical response used to identify material properties, while evaluating both through the same forward simulator.

C Experimental Details and Complete Results

C.1 Controlled Material Identification

The controlled material-identification experiment isolates physical parameter recovery from contact-state uncertainty by fixing the initial robot–object pose to its ground-truth value. The visual and tactile variants use the same prescribed material initialization and robot trajectory. We evaluate both single-parameter and multi-parameter identification settings.

C.1.1 Experimental Setup

The evaluated scenes are Apple, Banana, Star, and Bunny. Apple and Banana optimize Young's modulus E. Star jointly identifies Young's modulus E and Poisson's ratio ν, while Bunny jointly identifies Young's modulus E and yield stress σ_y.

Visual observations are captured by six fixed cameras arranged around the object at azimuth angles of 30°, 90°, 150°, 210°, 270°, and 330°, each approximately $0.5\,\text{m}$ from the observation center. Visual and tactile observations are sampled at the same time steps, at intervals of 10 frames.

Scene	Initial Guess	Optimized		Ground Truth
		Vision	Tactile	
Apple	$E = 500\,\mathrm{kPa}$	$E = 633.86\,\mathrm{kPa}$	$E = 995.67\,\mathrm{kPa}$	$E = 1000\,\mathrm{kPa}$
Banana	$E = 35\,\mathrm{kPa}$	$E = 43.11\,\mathrm{kPa}$	$E = 68.78\,\mathrm{kPa}$	$E = 70\,\mathrm{kPa}$
Star	$E = 60\,\mathrm{kPa},\ \nu = 0.4$	$E = 59.97\,\mathrm{kPa},\ \nu = 0.39$	$E = 32.42\,\mathrm{kPa},\ \nu = 0.265$	$E = 30\,\mathrm{kPa},\ \nu = 0.3$
Bunny	$E = 50\,\mathrm{kPa},\ \sigma_y = 2\,\mathrm{kPa}$	$E = 49.98\,\mathrm{kPa},\ \sigma_y = 2.16\,\mathrm{kPa}$	$E = 100.17\,\mathrm{kPa},\ \sigma_y = 5.00\,\mathrm{kPa}$	$E = 100\,\mathrm{kPa},\ \sigma_y = 5\,\mathrm{kPa}$

Table 2: Material-parameter identification from visual and tactile observations. Both modalities use the same material initialization, robot trajectory, ground-truth initial pose, and observation frames sampled at 10-frame intervals.

C.1.2 Visual and Tactile Optimization

The tactile formulation uses the pressure sequence at these shared observation frames while keeping the initial robot–object pose fixed. Material parameters are optimized using the differentiable forward model and adjoint-based material update described in the main paper.

For the visual baseline, reference object masks are obtained by localizing the object with GroundingDINO and segmenting it with SAM2. Candidate silhouettes are differentiably rendered from the simulated object geometry using nvdiffrast at the corresponding camera views and observation frames. The visual objective is the IoU loss between the rendered silhouettes and reference masks, without an RGB appearance term. Material parameters are optimized using L-BFGS-B. Optimization terminates when a gradient-based or loss-change convergence criterion is satisfied, or when it reaches 20 iterations or 80 objective-function evaluations, whichever occurs first.

C.1.3 Complete Results

Table 2 reports the initial guesses, optimized parameters, and ground-truth values for all four scenes. Across the reported single- and multi-parameter settings, tactile optimization yields estimates closer to the ground truth than the visual baseline.

C.2 Complete Inverse-Physics Evaluation with Pose Estimation

We next provide the complete implementation details and aggregate statistics for the unknown-contact experiments reported in Section 4.2. Across scenes, the task-constrained contact state contains translations, rotations, or their combination, while the physical parameter of primary interest is Young's modulus.

C.2.1 Observation Windows and Scene Configurations

Target and candidate runs within each scene use the same normal-pressure taxel layout, contact-patch asset, frame IDs, and time normalization. The full evaluation window is fixed for each scene. The material warm-up window is determined from the target sequence and ends 20 frames after the target's first contact, clipped to the fixed scene window. Neither window depends on candidate contact onset.

Pressure fitting uses the target-side normalization in Eq. (4). Pose fitting uses the `framewise_probability_moments_24b` contract with $N_{\mathrm{ref}} = 40$, $\tau = 0$, equal component weights, the time metric `squared_mean_event_frame_v1`, and the spatial metric `onset_aligned_continuous_pressure_center_v1`. Missing contact or pressure is marked invalid rather than assigned a fabricated moment. Material fitting uses `target_normalized_raw_mse`.

The eight scenes used in the main comparison are listed in Table 3. The scene names and ordering follow the main result table exactly.

C.2.2 Initialization Conditions

The main experiment uses deterministic coverage conditions rather than randomly sampled initialization seeds. For scenes with four translational starts, each pose initialization is crossed with the soft

Table 3: Ground-truth Young's modulus and fixed observation windows for the eight unknown-contact scenes used in the main experiment.

Scene	$E^\star$	Frames	Scene	$E^\star$	Frames
Apple	1.6 MPa	1–100	Star_r	5 MPa	1–70
Banana	0.2 MPa	31–110	Banana_r	0.2 MPa	31–120
Kong	0.5 MPa	1–120	Torus	0.75 MPa	160–260
Cup	2.5 MPa	20–140	Sphere	0.15 MPa	1–40

Table 4: Pose initializations used in the main experiment. Translations are in millimeters and rotations are in degrees.

Scene	Starts
Apple	$(-20,-4)$, $(-20,4)$, $(20,-4)$, $(20,4)$
Banana	$(-12,-10)$, $(-12,10)$, $(12,-10)$, $(12,10)$
Kong	$(-11,-14)$, $(-11,14)$, $(10,-14)$, $(10,14)$
Cup	$(-15,-15)$, $(-15,15)$, $(15,-15)$, $(15,15)$
Star_r	$\phi_y = -20^\circ,\ 20^\circ$
Banana_r	$(-12,-8,-20)$, $(-12,8,20)$, $(12,-8,20)$, $(12,8,-20)$
Torus	$(-20,-20)$, $(-20,20)$, $(20,-20)$, $(20,20)$
Sphere	$(-15,-6)$, $(-15,6)$, $(15,-6)$, $(15,6)$

and hard material entrances. Star_r contains two rotational starts, again crossed with the two material entrances.

The fixed pose initializations available in the current experiment records are listed in Table 4. Translations are reported in millimeters and rotations in degrees.

C.2.3 Compared Methods and Optimization Settings

We evaluate the following inference procedures.

- **Base**: keeps the initial contact state fixed and optimizes only the material parameters.
- **Raw-Pose**: follows the alternating refinement procedure but uses the full raw-pressure residual for pose estimation instead of the structured pose objective.
- **Joint Raw CMA**: jointly searches over the contact-state and material variables using CMA-ES and the raw-pressure objective.
- **Full w/o warmup**: uses the structured pose objective and alternating material refinement but removes the coarse material warmup.
- **Full w/o alt.**: uses the structured pose objective and the material warmup but removes alternation between pose and material updates.
- **Full**: uses the complete structured alternating procedure.

Their update schedules are summarized in Table 5.

The common optimizer settings are listed in Table 6.

The soft and hard material entrances are 100 kPa and 10 MPa for every scene. Base shares one soft trajectory between its two logical material sides; the remaining methods retain both entrances. CMA random streams are derived from the namespace, scene, start, material side, optimization block, and replicate.

C.2.4 Use of Ground Truth and Validity Criteria

Ground truth is used to generate synthetic targets, define frozen scene initialization conditions, construct controlled pose perturbations, and compute post-hoc errors. It is not used for candidate ranking, warmup selection, Armijo acceptance, online stopping, or CMA seed selection. Finite differences are used only for branch-local gradient checks.

Table 5: Optimization schedules used by the compared methods.

Method	Schedule
Base	warmup → continuous material refinement
Raw-Pose	Raw pose_1 → warmup_1 → material_1 → raw pose_2 → warmup_2 → material_2
Joint Raw CMA	One joint CMA-ES search over pose variables and $\log E$
Full w/o warmup	Pose_1 → material_1 → pose_2 → material_2
Full w/o alt.	Pose_1 → pose_2 → warmup → continuous material refinement
Full	Pose_1 → warmup_1 → material_1 → pose_2 → warmup_2 → material_2

Table 6: Optimizer configurations and evaluation budgets.

Configuration	Setting
Pose CMA-ES, each round	Population 6, 6 generations, 36 candidates
Joint Raw CMA	Population 8, 6 generations, 48 candidates
Material update limit	20 continuous updates per alternating block; 40 for Base and Full w/o alt.
L-BFGS	History 5, initial step 0.25, `normalize_dir=cold`, `grad_tol=1e-10`
Armijo backtracking	Constant 10^{-4}, at most 9 halvings, forward-only evaluation
CMA-ES implementation	`pycma==4.4.4`, `BoundTransform`, no restarts or built-in early stopping

For material refinement, simulator proposals must remain finite and numerically healthy and must strictly reduce the configured raw-pressure objective. Runs that do not produce a valid simulation result are excluded from the reported error statistics. These failures arise primarily when initial pose perturbations cause interpenetration between the object mesh and the PIPER gripper. Completed runs are retained regardless of their identification error.

Targets are generated three times and terminal states are replayed twice as part of the experiment validation procedure.

C.2.5 Additional Initialization-Sensitivity Variants

The `cma_random` variant changes only the CMA replicate from 0 to 1; pose starts, objectives, and evaluation budgets remain unchanged.

The `pose_random` variant perturbs the enabled translational initialization dimensions independently using

$$d \sim \text{TruncNormal}\left(0.5,\ 0.25^2;\ [0,1]\right)\ \text{mm}, \tag{50}$$

with a deterministic seed derived from the local experiment identifier, and updates each nonzero coordinate according to

$$x' = \text{sign}(x)\max(|x| - d, 0). \tag{51}$$

The sampled initialization is shared across methods and material sides, while rotations remain unchanged. This perturbation is a one-sided ground-truth-directed sensitivity test rather than zero-mean noise.

Torus instead uses one method-shared truncated-normal perturbation per start, with $\sigma = 0.5$ mm and rejection limits of $[-1, 1]$ mm for each translation component.

C.2.6 Complete Material and Pose Results

We provide the complete aggregate results underlying the compact comparison in the main paper. To characterize both typical performance and variability across different conditions, we report the mean, median, minimum, and maximum errors for each evaluated method. Material error is measured as

$$e_E = 100\left|\frac{\widehat{E}}{E^\star} - 1\right|\%. \tag{52}$$

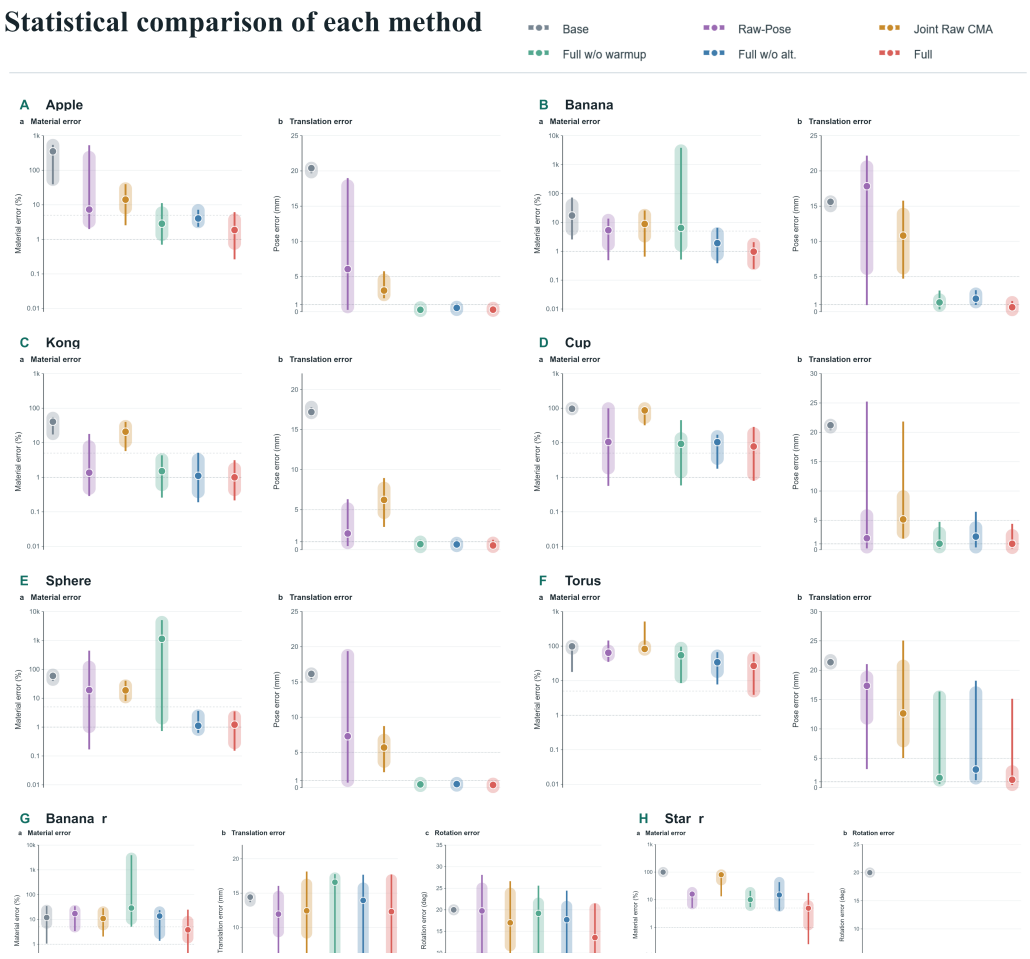


Figure 6: Statistical comparison of the evaluated methods across the eight unknown-contact scenes. Each scene reports material error together with translation or rotation error for its contact-state parameterization.

Translation error is the Euclidean error of the optimized translational contact variables in millimeters, while rotation error is the absolute world-Y angle error in degrees. For each scene and method, statistics are computed directly from all retained per-run results across the evaluated initialization conditions. The reported sample count N denotes the number of retained runs. These statistics describe completed runs rather than all attempted runs; exclusion criteria are given in Sec. C.2.4.

Figure 6 provides an overview of the error distributions across all eight scenes, while the following tables report the corresponding numerical statistics in full.

Across the complete set of conditions, Full consistently achieves low material and pose errors across most scenes. The gap between the mean and median in several configurations also reveals occasional large-error cases, particularly when material warmup is removed. Reporting both central-tendency and range statistics therefore exposes these failure cases rather than reducing each method to a single aggregate number.

C.2.7 Joint CMA-ES with Increased Search Budget

The computational cost of individual forward and backward evaluations varies substantially across scenes and initialization conditions, depending on the resulting contact configuration and simulation trajectory. Consequently, per-iteration runtime is not a representative metric for comparing the two

Table 7: Complete aggregate material and translation-error statistics for the six non-rotational scenes. Mean, median, minimum, and maximum are reported over complete conditions.

Scene	Method	N	Material error (%)				Translation error (mm)			
			Mean	Median	Min	Max	Mean	Median	Min	Max
Apple	Base	12	299.70	353.15	37.74	665.69	20.166	20.396	19.552	20.396
	Raw-Pose	24	153.33	7.31	0.66	702.52	9.123	6.069	**0.028**	20.396
	Joint Raw CMA	24	19.26	14.16	0.49	54.98	3.895	3.025	1.459	12.316
	Full w/o warmup	24	7.76	2.84	0.28	86.77	**0.376**	**0.287**	0.080	1.415
	Full w/o alt.	24	4.75	4.03	0.06	9.95	0.610	0.564	0.043	2.167
	Full	24	**2.58**	**1.87**	**0.05**	**7.97**	0.383	0.307	0.075	**1.162**
Banana	Base	12	26.52	17.24	1.71	73.93	15.386	15.620	14.678	15.620
	Raw-Pose	24	7.46	5.34	**0.01**	35.80	13.815	17.831	**0.151**	22.914
	Joint Raw CMA	24	11.40	8.80	**0.01**	44.38	10.568	10.818	3.608	18.654
	Full w/o warmup	23	1448.42	6.42	0.07	4919.57	1.442	1.333	**0.151**	3.774
	Full w/o alt.	24	2.72	1.91	0.06	10.41	1.943	1.865	0.477	3.563
	Full	23	**1.20**	**0.97**	**0.01**	**5.47**	**0.802**	**0.648**	0.159	**2.454**
Kong	Base	12	36.88	40.59	13.74	55.48	17.267	17.205	16.356	17.804
	Raw-Pose	24	5.25	1.36	0.06	22.08	2.869	2.032	0.128	7.843
	Joint Raw CMA	24	22.20	20.88	1.11	58.63	6.423	6.225	1.459	16.283
	Full w/o warmup	24	10.71	1.50	0.05	179.51	0.720	0.703	0.219	1.654
	Full w/o alt.	24	1.97	1.10	0.02	6.26	0.702	0.664	**0.038**	**1.524**
	Full	24	**1.43**	**1.00**	**0.01**	**5.50**	**0.669**	**0.535**	0.225	1.654
Cup	Base	12	97.08	96.29	94.86	99.60	20.981	21.213	20.032	21.213
	Raw-Pose	24	33.25	10.51	**0.14**	98.74	6.550	1.960	0.145	25.757
	Joint Raw CMA	24	85.17	86.47	20.28	405.85	7.879	5.171	0.848	25.642
	Full w/o warmup	24	16.92	9.27	0.36	**98.24**	2.024	1.027	**0.117**	**8.031**
	Full w/o alt.	24	**12.98**	10.44	0.36	99.02	2.594	2.242	0.142	**8.031**
	Full	24	14.19	**7.80**	0.30	99.53	**1.888**	**1.019**	0.153	10.095
Torus	Base	11	101.88	98.40	17.67	321.64	21.470	21.369	21.213	21.940
	Raw-Pose	24	161.88	64.48	25.69	1233.33	14.645	17.357	0.594	23.682
	Joint Raw CMA	24	167.25	82.83	37.42	517.95	14.382	12.646	1.983	25.368
	Full w/o warmup	24	55.04	54.41	3.09	151.76	6.188	1.671	**0.188**	19.500
	Full w/o alt.	24	35.86	34.04	2.97	**72.18**	8.634	3.092	0.497	19.911
	Full	23	**28.82**	**26.87**	**0.91**	86.24	**4.037**	**1.350**	0.329	**18.992**
Sphere	Base	12	53.72	59.27	30.25	64.11	15.934	16.155	15.073	16.155
	Raw-Pose	24	115.11	19.12	0.03	756.02	9.341	7.296	0.107	19.805
	Joint Raw CMA	24	2472.74	18.74	1.65	58753.36	6.168	5.700	1.439	18.085
	Full w/o warmup	24	2198.78	1136.44	0.25	6566.67	0.553	0.463	**0.069**	1.940
	Full w/o alt.	24	1.88	**1.11**	0.23	7.72	0.551	0.508	0.107	1.940
	Full	24	**1.43**	1.21	**0.01**	**4.22**	**0.444**	**0.382**	**0.069**	**1.673**

Table 8: Complete aggregate material and rotation-error statistics for Star_r. Rotation error is the absolute world-Y angle error in degrees.

Scene	Method	N	Material error (%)				Rotation error (deg.)			
			Mean	Median	Min	Max	Mean	Median	Min	Max
Star_r	Base	6	99.68	99.68	99.66	99.70	20.000	20.000	20.000	20.000
	Raw-Pose	12	14.30	15.99	1.00	35.94	0.350	**0.347**	0.025	0.785
	Joint Raw CMA	12	66.22	80.35	3.33	97.39	2.045	2.303	0.235	3.668
	Full w/o warmup	12	13.53	10.02	4.44	43.73	0.694	0.611	0.054	1.681
	Full w/o alt.	12	22.51	14.87	2.27	43.53	1.032	1.099	0.053	1.681
	Full	12	**6.04**	**4.92**	**0.11**	**18.85**	**0.348**	0.458	**0.021**	**0.680**

optimization strategies. We therefore compare their total GPU optimization time over complete identification runs.

For a conservative comparison, we restrict this experiment to scenes with only three optimized parameters. This setting favors the joint CMA-ES baseline, since derivative-free population-based search generally becomes increasingly expensive as the search dimension grows. We further increase its search budget from the 8×6 configuration used in the main experiments to 10×20, providing substantially more joint function evaluations.

This experiment tests whether the performance gap can be explained simply by insufficient joint-search computation. As shown in Table 11, even under this low-dimensional setting and with the enlarged search budget, joint CMA-ES requires more total GPU optimization time while still yielding

Table 9: Complete aggregate material and translation-error statistics for Banana_r.

Scene	Method	N	Material error (%)				Translation error (mm)			
			Mean	Median	Min	Max	Mean	Median	Min	Max
Banana_r	Base	12	15.58	11.90	0.23	35.80	14.184	14.422	13.518	**14.422**
	Raw-Pose	23	17.27	17.18	1.20	58.11	11.830	**11.947**	3.190	20.890
	Joint Raw CMA	22	12.95	10.84	0.32	**30.96**	12.474	12.435	1.652	22.972
	Full w/o warmup	20	1371.66	28.76	4.12	5078.29	11.975	16.598	1.337	18.414
	Full w/o alt.	22	13.76	13.64	**0.06**	58.11	11.124	13.944	0.976	18.032
	Full	21	**11.02**	**3.79**	0.16	58.58	**10.418**	12.302	**0.902**	19.008

Table 10: Complete aggregate rotation-error statistics for Banana_r.

Scene	Method	N	Mean	Median	Min	Max
Banana_r	Base	12	20.000	20.000	20.000	**20.000**
	Raw-Pose	23	16.647	19.753	0.450	29.340
	Joint Raw CMA	22	17.273	16.992	1.405	28.654
	Full w/o warmup	20	15.570	19.181	2.031	26.448
	Full w/o alt.	22	14.624	17.709	0.820	29.311
	Full	21	**12.939**	**13.565**	**0.153**	23.601

higher material errors than Full across all scenes. Contact-state errors are also generally higher. Overall, Full achieves better identification accuracy with lower total GPU cost, suggesting that increasing the joint-search budget alone does not eliminate the performance gap.

Figure 7 summarizes the same comparison across scenes. A ratio above one indicates a larger error or runtime for joint CMA-ES than for Full.

C.3 Real-Robot Experimental Setup

Our real-robot setup is shown in Figure 8. It consists of a PIPER robot equipped with two piezoresistive tactile sensors and three Intel RealSense RGB-D cameras (D405, D435i, and D455). The D405 is mounted near the end effector with a 15° tilted flange, while the other two cameras provide external views of the interaction. Each tactile sensor is mounted on the inner surface of a gripper finger and contains a 20×10 taxel array over an active sensing area of $3 \times 1.5\,\mathrm{cm}^2$.

Tactile measurements are recorded at 100 Hz, and robot and gripper states at 50 Hz. All three cameras capture 640×480 RGB-D observations at 15 fps. The visual observations are used only to reconstruct the object geometry and obtain a coarse initial pose. Physical-property identification uses the tactile pressure sequence recorded during a single probing interaction. The simulator follows the same physical time scale as the recorded robot interaction. For new-pose prediction, the object pose is manually specified as a simulation input, while the identified Young's modulus remains fixed.

Sensor limitation. We observe a failure mode for very soft objects. As the object deforms, the contact region can spread over a substantially larger area, resulting in weak local pressure measurements across the tactile array. When these signals fall below the effective sensitivity of the sensor, the tactile observations no longer provide sufficient information for reliable material identification.

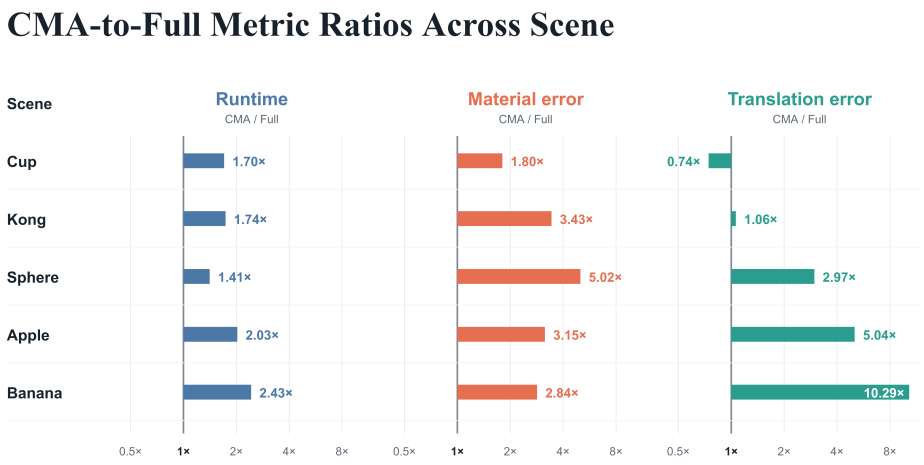


Figure 7: Scene-wise ratios of runtime, material error, and translation error for joint CMA-ES (10×20) relative to Full. A ratio of one indicates equal values.

Table 11: Comparison of joint CMA-ES (10×20) and Full on three-parameter material–translation estimation. For each object, per-run results from three initial-condition settings are combined before computing aggregate statistics.

Scene	Method	Material error (%)		Translation error (mm)		Time (min)
		Mean	Median	Mean	Median	Mean
Apple	CMA-ES (10×20)	5.682	5.718	2.070	1.653	56.601
	Full	**2.582**	**1.865**	**0.383**	**0.307**	**27.844**
Banana	CMA-ES (10×20)	4.436	1.746	8.021	5.574	102.126
	Full	**1.198**	**0.970**	**0.802**	**0.648**	**40.011**
Kong	CMA-ES (10×20)	5.001	3.294	1.336	0.540	49.078
	Full	**1.427**	**1.000**	**0.669**	**0.536**	**32.716**
Cup	CMA-ES (10×20)	24.529	13.768	1.934	1.035	80.317
	Full	**14.189**	**7.805**	**1.888**	**1.019**	**46.692**
Sphere	CMA-ES (10×20)	1263.047	5.309	2.472	1.288	46.086
	Full	**1.428**	**1.215**	**0.444**	**0.382**	**20.067**

Each CMA-ES setting contributes eight completed trajectories. Full uses the same nominal initialization coverage, with one incomplete Banana trajectory excluded. Means and medians are computed directly from the retained per-trajectory records across the three settings. Runtime reports accumulated GPU evaluation time rather than wall-clock time.

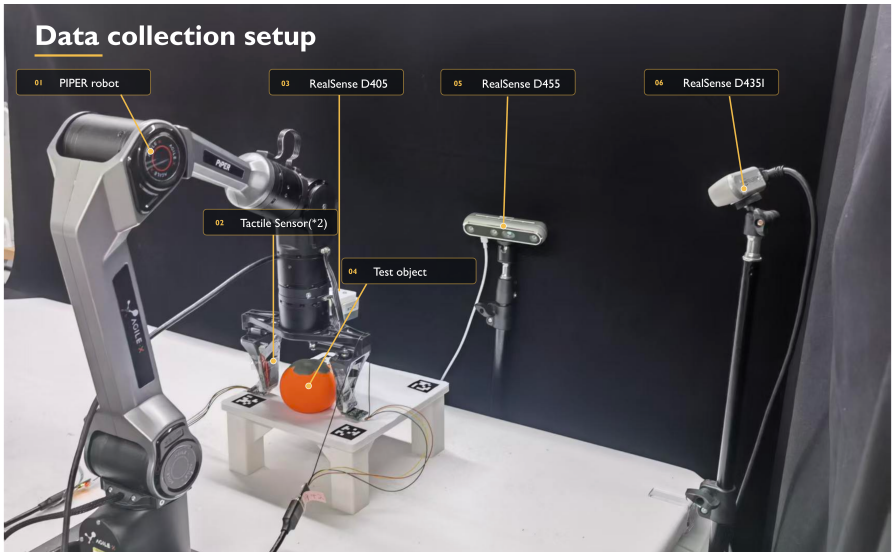


Figure 8: Real-robot experimental setup. Two tactile sensors are mounted on the gripper fingers, while three RGB-D cameras provide visual observations for object reconstruction and pose initialization.