NoContactNoWorries: Estimating Contact through Vision & Proprioception for In-Hand Dexterous Manipulation

Soham Patil[1] · Avirup Das[2] · Sourabh Bhosale[1] · Spandan Roy[1]
[1]Robotics Research Center, IIIT Hyderabad [2]Dept. of Computer Science, University of Manchester
https://soham2560.github.io/no-contact-no-worries

IROS 2026 PITTSBURGH

Overview

Problem: In-hand manipulation needs finger, object contact; tactile sensors are costly, fragile, hard to scale.

Idea: Infer binary fingertip contact from ego-centric RGB-D + proprioception, no tactile hardware.

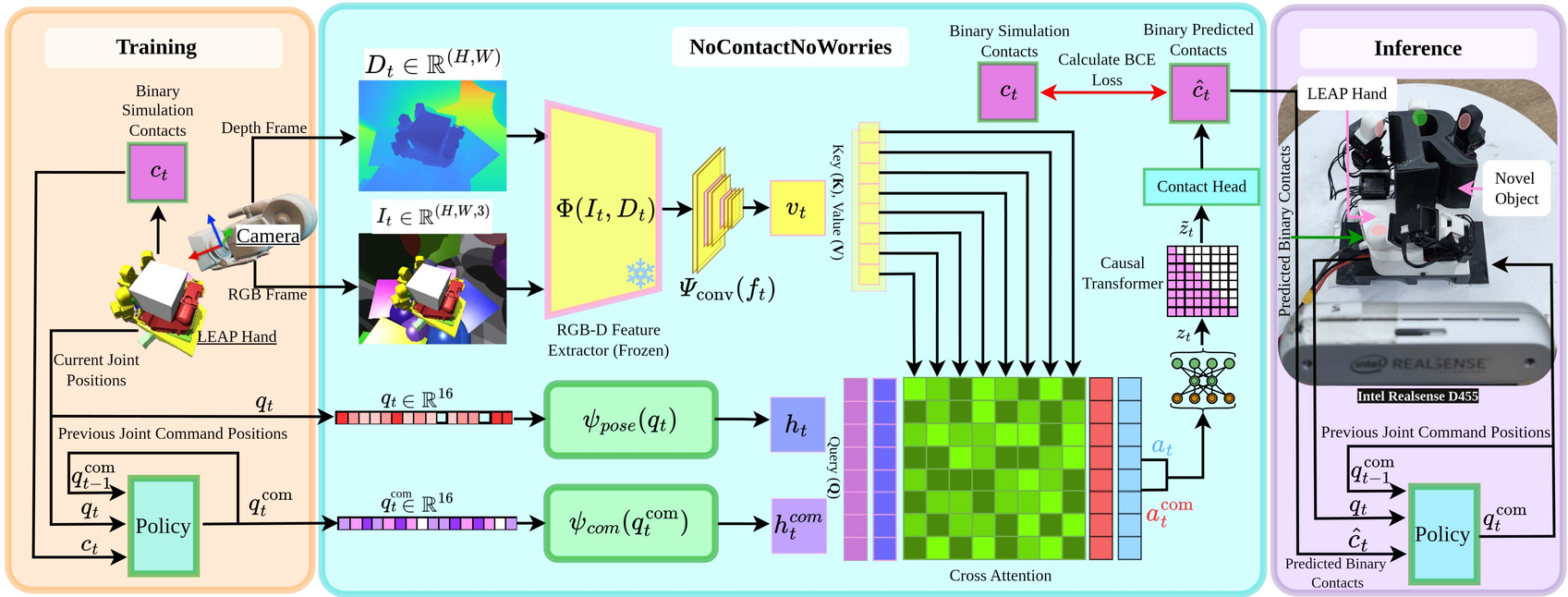
Method: A transformer fuses vision with joint states and reasons over time → per-fingertip contact $\hat{c}_t$

Use: Predicted contact replaces oracle tactile in RL policies; sim-to-real, no retraining.

Contributions

Pseudo-tactile signal from RGB-D + proprioception, no tactile hardware.
Pose-conditioned cross-attention: reactive q_t + anticipatory q_t^{com}.
Causal temporal reasoning: motion cues, not single-frame proximity.
Drop-in surrogate: sim-to-real on the LEAP hand; generalizes to novel objects.

Method: NoContactNoWorries



Contact Prediction

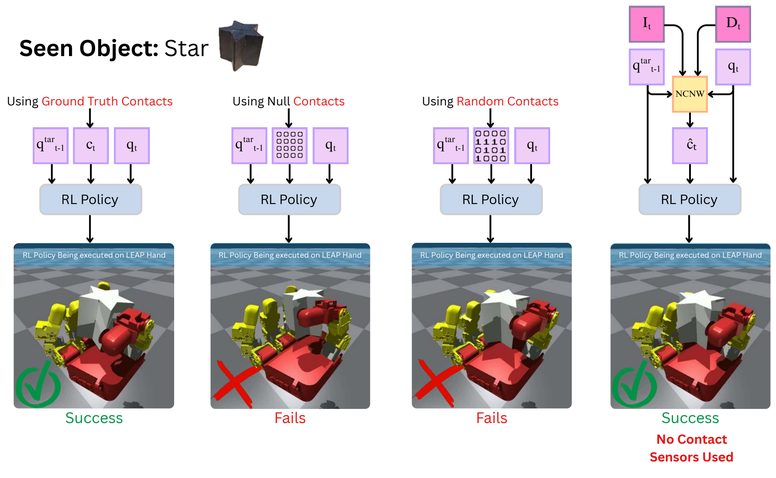
0.93 / 0.85 F1 — visible / occl.	**0.79 → 0.51** vision-only, occluded
~61% frames self-occluded	**8 ms** inference (20 Hz)

Model	Vis.	Occl.
Full (Proposed)	**0.93**	**0.85**
Pose-only (best)	0.75	0.81
Vision-only	0.79	**0.51**
Kinematic depth	0.81	0.54

Robust under occlusion; vision-only & depth baselines collapse.

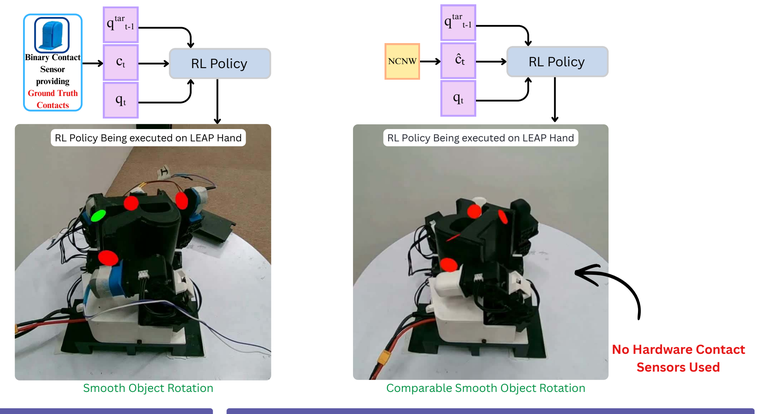
Simulation Results: In-Hand Rotation

Seen Object: Star

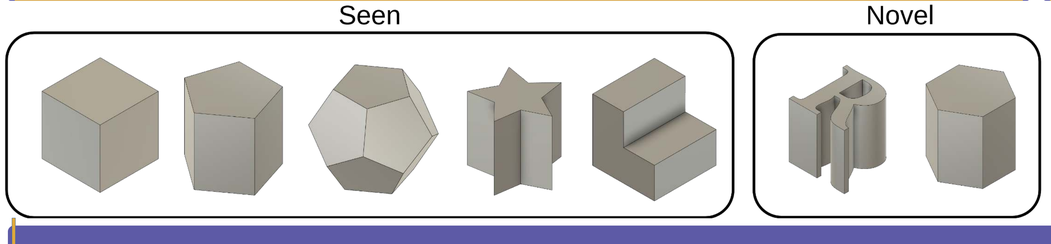


Hardware Results: In-Hand Rotation

Novel Object: Letter R



Object Set (Seen / Novel)

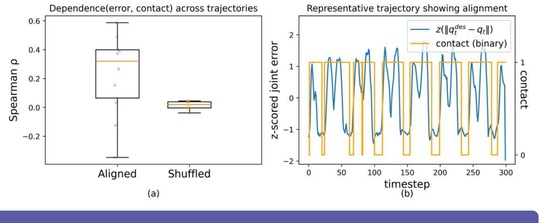


Downstream: In-Hand Rotation (CRA)

Object	No-Contact	Oracle	Ours
Cuboid	19.46	19.10	**20.24**
Dodecahedron	18.95	18.43	**19.40**
Star	13.25	14.95	14.37

CRA (rad), real hand, 10 trials. Ours ≥ Oracle on convex; small gap on non-convex (star).

Pose → Contact



Takeaway. Pose-conditioned cross-attention + causal temporal fusion turn RGB-D and proprioception into a drop-in contact signal, near-oracle in-hand rotation on real hardware, no tactile sensors, no retraining.