T-FRAME: TACTILE FORCE-AWARE REPRESENTATION FOR ACTION-CONDITIONED MANIPULABILITY EVOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reliable dexterous fastening requires repeated contact renewal until a part reaches its seat. Maximizing the current turn can compromise release and recontact, making immediate progress a poor guide to sustained operation. We introduce T-FRAME, which learns action-conditioned contact events, loads, and remaining manipulation resources alongside continuation value. Physical outcomes supervise action consequences, while preferences train choices among alternatives from the same contact state. Uncertain contact and relative-pressure evidence from human video shapes these preferences; tactile-conditioned GPU search generates robot-native trajectories evaluated through physical rollouts. Selected trajectories and task reinforcement learning train a recurrent policy that adjusts finger and wrist commands using live touch, without online search. A finite-candidate analysis separates action coverage, teacher alignment, and selection error. Across simulated fastening tasks with different head shapes, diameters, and stable or initially loose engagement, T-FRAME achieves near-complete insertion progress and shorter successful completion times than the adapted baselines. Ablations link human evidence to execution efficiency, directional touch to reduced backdrive, and current feedback to sustained completion. These results support learning contact consequences as a basis for reliable, efficient long-horizon fastening.

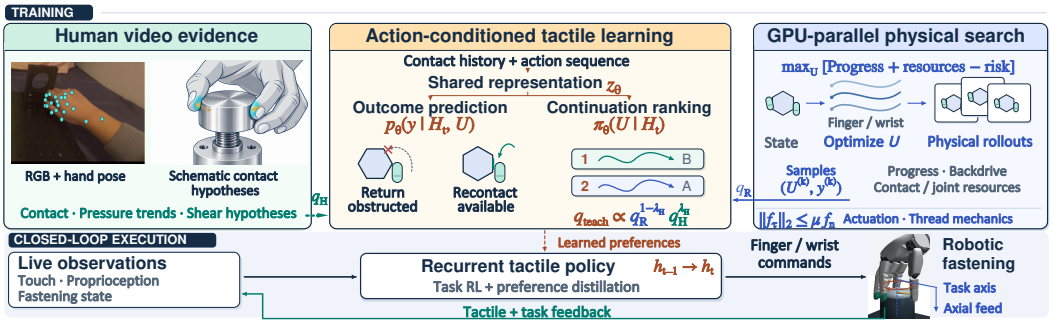

Figure 1: T-FRAME overview. Human contact evidence and physical outcomes of GPU-generated candidates supervise action-conditioned prediction and preference learning. Selected trajectories train a recurrent policy that uses live tactile and proprioceptive feedback to renew contact and advance the fastener through seating. Video processing and trajectory search are training-time components.

1 INTRODUCTION

Threaded fastening requires repeated traction, finger repositioning, and axial feed until a part reaches its seat. A dexterous hand offers a common interface for graspable heads of different shapes and sizes, but industrial utility requires more than producing a turn: reliable completion through contact changes and efficient progress without repeated reversal.

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.

Dexterous rotation and fastening can be learned through reinforcement learning, human guidance, and motor pretraining (Hsieh et al., 2026; Chen et al., 2026; Lum et al., 2026). Consider two drive actions from the same contact state. A larger turn may consume joint travel or leave residual compression that obstructs return, whereas a smaller turn may preserve a productive next grasp. Return-based learning can represent this tradeoff; the challenge is learning its dependence on contact geometry and load from finite interaction. Similar hand postures can correspond to weak contact, effective traction, or loaded slip. Changes in head shape and resistance further alter the consequences of the same command. Task return alone does not explicitly supervise these distinctions.

Our central idea is to supervise *action-conditioned tactile force–motion evolution* alongside continuation value. Contact events, loads, and remaining resources explain the consequences behind a score; preferences among same-state alternatives connect it to selection. Matched value-only learning tests whether this structured supervision improves decisions.

Human video contributes uncertain contact cues; GPU-parallel search supplies physically evaluated robot actions (Figure 1). Human evidence shapes preferences, not physical outcome labels. Selected trajectories and task interaction train a recurrent policy driven by online touch.

Our contributions are threefold:

- **Learning contact consequences for action selection.** Joint outcome and preference supervision shapes a candidate-conditioned representation beyond scalar continuation value.
- **Physics-based supervision at scale.** Tactile-conditioned search explores uncertain contact mechanics; finite-candidate analysis separates coverage, teacher alignment, and selection error.
- **Transferring contact evidence into feedback control.** Uncertainty-weighted human evidence guides trajectory distillation and task learning without requiring video or search at execution time.

Experiments test completion, efficiency, and generalization under stable and initially loose engagement, together with dependence on live touch.

2 Related Work

Video contact and human guidance. EgoPHI estimates mesh-level contact and force from images and known geometry (Ilic et al., 2026); EgoTac learns temporal contact and scalar pressure (Zhang et al., 2026). C2Dex preserves canonical object contacts in retargeting (Ren et al., 2026), while DexCanvas supplies simulation-derived force profiles (Xu et al., 2025). DexMani transfers rotational manipulability evolution (Chen et al., 2026); CHORD transfers contact-wrench guidance (Zhu et al., 2026). Rather than treating human force estimates as unique targets, we use uncertain cues to shape preferences and physical robot outcomes to supervise realized contact changes and remaining resources.

Tactile control and industrial manipulation. RotateIt and AnyRotate use rich sensory feedback for sustained in-hand rotation (Qi et al., 2023; Yang et al., 2025). DexScrew learns rotational gaits in simulation and uses them to collect real multisensory demonstrations for policy learning (Hsieh et al., 2026). SimToolReal learns object-centric goal tracking across procedurally generated tool-like objects (Kedia et al., 2026); Play2Perfect adapts task-agnostic play to precise assembly (Lum et al., 2026). These motivate comparisons with goal following and motor pretraining. Spatially anchored or physics-grounded tactile features (Huang et al., 2025; Pan et al., 2026) and tactile pretraining (Higuera et al., 2025; Feng et al., 2025; Liu et al., 2025) address sensing and representation; our focus is their action-conditioned evolution for continuation-aware decisions. Factory supplies contact-rich assembly simulation (Narang et al., 2022); we isolate repeated fastening after correct thread registration.

Physical models for learning and search. Control-oriented representations (Zhang et al., 2021) and decision-focused learning (Mandi et al., 2022) prioritize downstream behavior; TD-MPC2 plans with learned latent dynamics (Hansen et al., 2024). CuRobo and cuNRTO parallelize motion generation and robust trajectory optimization (Sundaralingam et al., 2023; Wang et al., 2026); differentiable convex layers and DiffTactile connect optimization and contact simulation to learning (Agrawal

et al., 2019; Si et al., 2024). We instead use local mechanics and physical rollouts during training to supervise contact outcomes and preferences, then distill decisions into recurrent control. This avoids online search and differentiation through contact switches; matched candidate sets separate representation quality from proposal quality.

3 Problem Formulation

We consider direct-finger insertion of a correctly started headed fastener into a fixed female-threaded base. The head and shank form one rigid body. Task parameters ψ specify external head geometry, grasp diameter, thread lead, handedness, target depth, support, and resistance. Grasp diameter describes the head's outer envelope and is distinct from nominal thread diameter.

Let $\hat{a}$ be the fixed fixture-axis direction in world coordinates, defining the positive axial coordinate z, and let θ be unwrapped rotation under the right-hand rule about $\hat{a}$. For pitch p, start count n_s, and thread handedness $s_{\mathrm{th}} \in \{-1, 1\}$, define the signed lead $\ell = s_{\mathrm{th}} n_s p$. The constrained screw coordinate satisfies

$$z - z_0 = \frac{\ell}{2\pi}(\theta - \theta_0), \qquad t_\psi = d \begin{bmatrix} (\ell/2\pi)\hat{a} \\ \hat{a} \end{bmatrix}, \tag{1}$$

where $d \in \{-1, 1\}$ is the commanded rotation direction. Thus positive commanded angular progress $\xi = d(\theta - \theta_0)$ produces signed axial travel $\ell d\xi/(2\pi)$. Reported insertion depth is $z_{\mathrm{ins}} = s_{\mathrm{ins}}(z - z_0)$, where the fixture specifies $s_{\mathrm{ins}} \in \{-1, 1\}$ and insertion commands satisfy $s_{\mathrm{ins}} \ell d > 0$. Handedness therefore does not by itself specify the command direction. Mechanics equations use radians; reported rotation and backdrive use $^\circ$. The twist t_ψ is evaluated at a point on the fixture axis, per unit angular progress, not as a unit vector mixing meters and radians.

We distinguish *stable engagement*, in which the correctly registered fastener is axis-aligned, from *initially loose engagement*, in which it can translate laterally and tilt within its support. Neither setting starts at the final seat. Both retain the correct thread start and the helical feed relation; individual thread-flank contact and cross-threading are outside this model. Success requires complete angular progress and target insertion, followed by the specified torque and low-speed seating hold.

At time t, observations contain proprioception, task/object state, and continuous tactile history, encoded as h_t. Contact features use object coordinates. Forces act on the object with compressive normals pointing into it; wrenches use the center of mass unless explicitly referred to the fixture axis (Appendix A.1). A training candidate $u_j \in \mathcal{U}_t$ is a bounded native command sequence $a^{(j)}_{0:H_a-1}$ of joint targets and permitted wrist residuals, not object motion. The deployed recurrent policy outputs one command per step rather than a candidate identity (Appendix A.6). The objective is reliable finite-depth completion with low execution time and backdrive.

Human RGB sequences provide additional training evidence without target-task tactile measurements. Even if the contact geometry and net wrench w are known, a feasible force solution has the form $f = G^\dagger w + N\alpha$, with $GN = 0$; internal forces remain undetermined. Human supervision therefore constrains supported contact events, motion, and pressure trends, while robot touch supplies execution-side load.

4 T-Frame

Figure 1 links human contact evidence, physical candidate evaluation, and action-conditioned learning to recurrent execution.

4.1 Temporal contact evidence from human video

Spatial hand–object correspondence locates candidate contacts; temporal observations provide onset/release events, relative surface motion, and pressure trends. Inferred cues remain distinct from supplied pose annotations. Pressure scores are relative unless calibrated with force scale and contact area.

A persistent contact carries a tangential displacement hypothesis. Let v_t^τ be hand-surface velocity minus object-surface velocity, projected into the local tangent plane, and let $f_t^{\tau,o}$ act on the object.

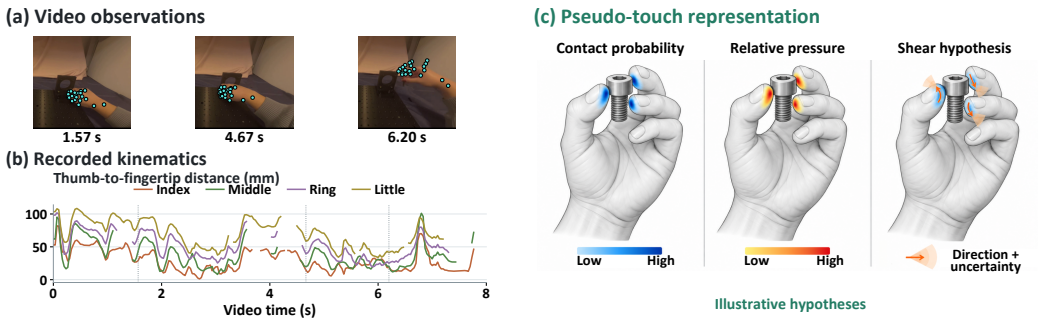

Figure 2: Temporal observations and contact representation. (a) EgoDex frames with dataset hand-pose observations (Hoque et al., 2026). (b) Thumb-to-fingertip distances from those annotations; dotted lines locate the frames. (c) Illustrative contact, pressure, and shear hypotheses, not predictions on this clip.

Then

$$\widetilde{\delta}^{\tau}_{t+1} = \mathcal{P}_t(\delta^{\tau}_t + \Delta t\, v^{\tau}_t), \qquad f^{\tau,o}_{t+1} = \Pi_{\|f\|\le\mu f^{n}_{t+1}}\left(k^{\tau}\widetilde{\delta}^{\tau}_{t+1} + c^{\tau}\mathcal{P}_t v^{\tau}_t\right). \tag{2}$$

Here $\mathcal{P}_t$ transports history and velocity to the updated tangent plane. Projection corrects elastic memory; release resets it (Appendix A.5). Pose, friction, and compliance uncertainty yield multiple hypotheses whose disagreement limits cue influence.

For robot transition $Y(u)$ and human window o^H, P_Ω retains comparable event, motion, and pressure-trend channels. Contacts match by object-relative position and surface orientation, scaled by head radius/height; unsupported matches are omitted. Monotone alignment $\mathcal{W}$ accommodates execution speed. With reliability weight $\omega_{kc} \in [0,1]$ for time k and channel c, agreement is

$$D_\Omega(Y, o^H) = \frac{\sum_{k,c}\omega_{kc}\,\ell_c\big([P_\Omega Y]_{k,c}, [P_\Omega o^H]_{\mathcal{W}(k),c}\big)}{\sum_{k,c}\omega_{kc} + \epsilon}. \tag{3}$$

Dimensionless losses compare events, motion directions, and pressure changes. Training-bank windows are selected by task direction and observed prefix, never future robot outcomes. This gives $U_H(u) = -\mathbb{E}[D_\Omega(Y(u), o^H)]$, where the expectation averages the retained hypotheses, without fixing finger identity.

Let q_R be the physical-outcome preference defined below. Human agreement modifies its training target through

$$q_H(u) \propto e^{U_H(u)/T_{\text{vid}}}, \quad T_{\text{vid}} = \tau_0 + \kappa\bar{\sigma}_H, \quad q_{\text{teach}}(u) \propto q_R(u)^{1-\lambda_H} q_H(u)^{\lambda_H}. \tag{4}$$

Hypothesis disagreement $\bar{\sigma}_H$ softens the human preference, while $\lambda_H \in [0,1)$ scales with usable evidence. If no comparison is supported, $\lambda_H = 0$ recovers q_R exactly. This interface changes relative supervision, not the measured progress, force, or failure of a robot rollout.

4.2 Tactile-conditioned GPU candidate optimization

Search starts from the current control state and measured normal/shear loads. Its variables are native action sequences around a policy or recorded robot motion warm start. Let H_a denote the candidate horizon and ζ a sampled mechanical scenario. The local objective is

$$\max_{a_{0:H_a-1}} \; \mathbb{E}_\zeta[\bar{\Delta}\xi + \lambda_R R_{H_a}] - \lambda_C\,\mathrm{CVaR}_\beta(C_\zeta) - \lambda_a \sum_{k=0}^{H_a-2} \|a_{k+1} - a_k\|^2. \tag{5}$$

Bars denote fixed-scale normalization of progress and resources; native actions are dimensionless. The nonnegative normalized cost C_ζ combines failures and excessive loads, with upper-tail CVaR emphasizing adverse scenarios. Candidates explore actions; scenarios sample mechanical uncertainty.

The terminal score $R = w_R^\top \bar{r}$ summarizes resources available for the next operation. We distinguish joint reserve, local task motion, and current-load torque capability:

$$r(s) = \big[r_q, m_\psi, r_\tau, r_{\text{release}}\big], \qquad \Delta r(u) = r(s^+) - r(s). \tag{6}$$

Here r_q is normalized joint-limit margin; m_ψ bounds a local screw-motion step under contact Jacobians and joint limits; r_τ compares a current-load torque upper bound with required torque; and r_{release} measures the fraction of contacts removable while retaining that capacity (Appendix A.4). Their evolution describes what an action leaves behind; physical continuation evaluates actual recontact and subsequent progress.

Local mechanics include unilateral contact, friction, and fixture-conditioned wrench balance:

$$f_i^n \geq 0, \quad \|f_i^\tau\|_2 \leq \mu_i f_i^n, \quad (q, \dot{q}, \tau) \in \mathcal{A}_{\text{act}}, \quad Gf + B_\psi \lambda_{\text{fix}} + w_g = M\dot{V} + h_{\text{dyn}} + w_{\text{res}}. \tag{7}$$

Here $\mathcal{A}_{\text{act}}$ is actuator admissibility and $B_\psi \lambda_{\text{fix}}$ is the fixture reaction. Wrenches share the center-of-mass reference; h_{dyn} contains velocity-dependent dynamics. Actual resistance is $-w_{\text{res}}$, with $V^\top w_{\text{res}} \geq 0$ for passive running resistance. Stable reactions are reciprocal to screw motion; loose support reacts through clearance contacts. The compliant proposal model is local, not a feasibility certificate: physical rollouts resolve saturation, support changes, and new contacts (Appendix A).

Bounded temporal knot offsets perturb a warm-start trajectory. GPU-parallel candidate/scenario evaluation retains the incumbent and updates the sampling distribution from elites. The compliant model predicts loads and projects shear onto the friction disk, separately from action clipping (Appendix A.3).

Retained candidates execute from the same simulator/controller state and contact history. A common frozen policy π_b continues each branch to $H \geq H_a$ rollout steps, giving all candidates equal time. Physical execution supplies labels, including failures, rather than search scores.

The role of GPU search is to construct alternative contact transitions from the same physical state, not merely to replay a nominal gait faster. Batching spans both finger/wrist trajectories and mechanical scenarios; friction and actuation constraints screen proposals, while physical continuation determines whether they preserve productive recontact. This separates generating useful training alternatives from learning to select them, with no search required at execution time.

4.3 Learning action consequences and preferences

The encoder pools contact position, normal, normal/shear load, relative slip, age, and onset/release indicators, then recurrently fuses them with proprioception and task observations: hand-relative object pose/velocity, insertion, and task axis. This preserves continuous load and geometry under fixed training scales (Appendix A.6). Candidate features encode proposed controls, not future measured outcomes.

Candidate-conditioned attention predicts utility and physical outcomes:

$$(\widehat{v}_\theta, \widehat{P}_\theta(\chi \mid h_t, u, \psi)) = F_\theta(h_t, u, \psi), \qquad \chi = (\Delta\xi, B, D, C, e^+, c^+, \Delta r, r^+). \tag{8}$$

The terminal event e^+ describes contact transitions; c^+ summarizes load, active contacts, and task torque. Torque includes normal/shear forces and any modeled patch moment about the fixture axis, not merely the center of mass (Appendix A.1). The resource targets Δr and r^+ distinguish change from remaining capacity. These outcomes separate actions with similar progress but different residual load or contact renewal; missing measurements do not remove failure labels (Appendix A.6).

To define the utility target, distinguish cumulative backdrive $B = \sum_k [-(\xi_{k+1} - \xi_k)]_+$ from maximum drawdown $D = \max_{j \leq k}(\xi_j - \xi_k)$. The first measures total reversal; the second measures the largest loss of attained progress. Under the common continuation policy π_b, physical branches define

$$\begin{aligned} Q_H(u; \pi_b) &= \mathbb{E}_\zeta \left[\sum_{k=0}^{H-1} \gamma^k \bar{\Delta}\xi_k - \lambda_D \bar{D}_H + \lambda_R R_H \right] - \lambda_C \, \mathrm{CVaR}_\beta(C_\zeta), \\ q_R(u) &\propto e^{Q_H(u;\pi_b)/\tau_R}. \end{aligned} \tag{9}$$

Progress and drawdown share an angular scale; normalized resources and costs make Q_H and its temperatures dimensionless. This target is conditioned on π_b, not an intrinsic success probability.

The predicted utilities supply ranking logits through $\pi_\theta = \mathrm{softmax}_{u \in \mathcal{U}_t}(\widehat{v}_\theta / \tau_s)$. Utility and outcome prediction share the candidate-conditioned representation. Their loss combines physical supervision with relative preference supervision:

$$\mathcal{L}_{\text{repr}} = \lambda_v \ell_v(\widehat{v}_\theta, Q_H) + \lambda_{\text{tr}} \mathbb{E}[-\log \widehat{P}_\theta(\chi \mid h_t, u, \psi)] + \lambda_{\text{pref}} \, \mathrm{CE}(\mathrm{sg}[q_{\text{teach}}], \pi_\theta). \tag{10}$$

Utility regression anchors continuation quality; outcome prediction shapes the shared contact representation; preference CE trains relative choices. Human guidance can alter ordering, so the score is not an unbiased physical return estimate.

For a fixed utility $Q = Q_H(\cdot; \pi_b)$ and admissible action class, the quality of a selected candidate separates into two terms:

$$V^*_{\mathrm{adm}} - Q(\hat{u}) = \underbrace{V^*_{\mathrm{adm}} - \max_{u \in \mathcal{U}_t} Q(u)}_{\epsilon_{\mathrm{support}}} + \underbrace{\max_{u \in \mathcal{U}_t} Q(u) - Q(\hat{u})}_{\epsilon_{\mathrm{rank}}}. \tag{11}$$

Search targets the quality of available choices; learning targets how well those choices are used. Best-candidate utility measures proposal quality, whereas the gap to selected-candidate utility measures selection error (Appendix D.5).

The mixed teacher in Eq. (4) is Gibbs with score $S = (1 - \lambda_H)Q + \lambda_H \tau_R U_H / T_{\mathrm{vid}}$. For a selection distribution p over the same N candidates, let $\Delta_S = \max S - \min S$ and $\epsilon_H = \mathrm{range}(Q - S)$. Then

$$V^*_{\mathrm{adm}} - \mathbb{E}_p Q \le \epsilon_{\mathrm{support}} + \epsilon_H + \tau_R \log N + \Delta_S \sqrt{\tfrac{1}{2} \mathrm{KL}(q_{\mathrm{teach}} \| p)}. \tag{12}$$

The result follows from Gibbs optimality and Pinsker's inequality (Appendix B). For $p = \pi_\theta$, cross-entropy training controls the final KL term up to teacher entropy. The other terms expose errors that ranking alone cannot remove: absent useful actions and a teacher preference that disagrees with robot utility. This finite-set analysis concerns expected selection, not a guarantee for greedy selection or the distilled controller.

4.4 Recurrent tactile execution

The execution policy $\pi_\phi(a_t \mid \mathcal{H}_t, \psi)$ maps causal sensor history $\mathcal{H}_t$ to continuous native commands. It has its own recurrent state and does not classify state-specific trajectory identities. For each training parent, the learned score selects u^*; its physical rollout supplies observation–action pairs for a distillation set $\mathcal{D}_{\mathrm{selected}}$. The actor objective combines task RL and action supervision (Rusu et al., 2016):

$$\begin{aligned} u^* &= \arg\max_{u \in \mathcal{U}_t} \widehat{v}_\theta(h_t, u, \psi), \\ \mathcal{L}_{\mathrm{actor}} &= \mathcal{L}_{\mathrm{PPO}} + \lambda_{\mathrm{dist}} \mathbb{E}_{(\mathcal{H}, a^*, \psi) \sim \mathcal{D}_{\mathrm{selected}}} [\ell_A(\mu_\phi(\mathcal{H}, \psi), a^*)]. \end{aligned} \tag{13}$$

Selecting one continuation per parent avoids averaging incompatible drive and unload actions. The teacher is fixed during each actor update stage; supervised initialization precedes task RL, with distillation retained as a regularizer (Appendix A.6). Task rewards favor signed progress and completion while penalizing reversal, drift, and effort.

Fresh normal/shear loads, slip, and proprioception update joint/wrist commands at every step, distinguishing weak contact from loaded slip during drive and residual shear during return. The known lead provides an axial reference, $z^*_{W,t} = z_{W,0} + \ell(\theta_t - \theta_0)/(2\pi) + \delta z_t^{\mathrm{touch}}$, where wrist height and residual are signed coordinates along $\hat{a}$, not the always-positive insertion depth z_{ins}; axis-centered rotation assists traction, while bounded transverse and tilt residuals accommodate initial play. Object-relative coordinates account for nominal feed; depth and resistance remain observable as contact dynamics change. Execution never resets the hand or fastener. Paired feedback interventions test whether learned corrections support continued insertion.

5 Experiments

5.1 Tasks and evaluation protocol

Sharpa Wave inserts correctly started fasteners through 6 mm over three revolutions, then maintains a low-speed seat. Four head shapes—round, octagonal, hexagonal, and square—and D64/D80/D96 grasp envelopes yield twelve stable conditions; four initially loose D80 shapes complete the 16-condition matrix, each evaluated over 100 trials. Thread lead and nominal resistance remain fixed (Appendix D). Constrained helical feed measures insertion progress rather than learned thread kinematics. Policies are frozen before testing; normalization and human-evidence selection use

Table 1: Fastening performance and core ablations: completion progress PR (%), completion time T (s), and backdrive B (°). T-FRAME uses 100 evaluation trials per task condition; baselines are adapted to the target tasks. SimToolReal has no seated successes, so T is undefined (–).

	Stable engagement			Initially loose		
Method	PR ↑	T ↓	B ↓	PR ↑	T ↓	B ↓
Play2Perfect (adapted)	37.2	114.5	48.8	31.4	120.8	87.3
SimToolReal (adapted)	15.4	–	7.9	25.6	–	18.8
DexScrew (adapted)	90.2	87.7	45.5	92.4	105.9	66.2
T-FRAME	99.3	58.9	35.6	99.0	63.8	44.1
Core ablations						
Without human evidence	98.7	70.2	38.8	98.1	69.4	45.0
Normal-only touch	97.5	59.3	60.5	95.9	63.9	80.2
Stale tactile feedback	85.5	77.1	48.2	73.6	90.5	114.8
Value-only supervision	81.2	73.4	83.9	75.7	101.5	109.6

training data only. Comparisons share test instances, a 450 s deadline, and physical limits. Controlled ablations match sensing and training budgets except for the stated intervention; external policies retain method-specific inputs (Appendix D).

5.2 BASELINES AND METRICS

We adapt Play2Perfect's play-to-assembly policy (Lum et al., 2026), SimToolReal's object-centric goal tracking (Kedia et al., 2026), and DexScrew's fastening skills with tactile history (Hsieh et al., 2026). All scores come from target-task evaluations, not the original publications. Four ablations test human evidence, directional touch, current feedback, and joint outcome–preference supervision.

Completion and efficiency. Completion progress (PR) is terminal directed net rotation divided by the required rotation, clipped to 0–100%. Seated success (SR) additionally requires complete progress, target depth, and the seating hold. Completion time T averages successful episodes and is undefined without successes. Cumulative backdrive B sums actual reverse object rotation, excluding resets. PR and B include failures. Table 1 averages evaluation trials within each task condition, then weights task conditions equally within each engagement regime; Figure 4 reports per-condition SR. Definitions appear in Appendix D.

5.3 LONG-HORIZON COMPLETION AND GENERALIZATION

Table 1 compares sustained progress, completion time, and backdrive under stable and initially loose engagement.

Play2Perfect initiates rotation but, in inspected rollouts, loses a usable configuration for subsequent strokes. SimToolReal instead stalls early: low backdrive accompanies low progress and zero seated successes, rather than indicating efficient control. Unfavorable geometry or opposing contact efforts are plausible explanations, not conclusions from the aggregate means. These comparisons concern adaptation to our tasks, not native-task capability. Fastening-specific DexScrew achieves substantially greater progress, but T-FRAME improves both stable/loose PR (99.3%/99.0% versus 90.2%/92.4%) and successful completion time. DexScrew's longer execution and greater backdrive are consistent with interrupted or less efficient traction.

Seated success ranges from 97% to 100% across conditions, totaling 1182/1200 (98.5%) for stable and 396/400 (99.0%) for initially loose engagement. Figures 3 and 4 show contact renewal, condition-level success, and insertion trajectories. The illustrative trajectories in Figures 3–5 come from separate reference executions (Appendix E). Residual failures involve seating transients and corner entrapment (Section 6); additional sequences appear in Appendix E.

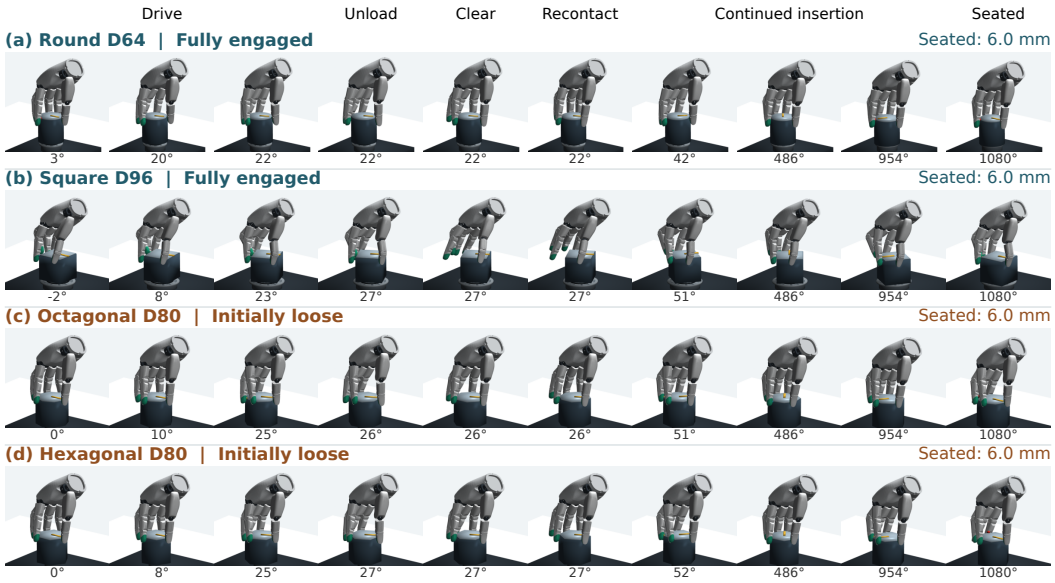

Figure 3: Contact renewal through seating: stable round D64 and square D96, then initially loose octagonal and hexagonal D80. Each trajectory has ten chronological, nonuniformly spaced frames labeled with measured cumulative rotation; the row header reports final insertion depth.

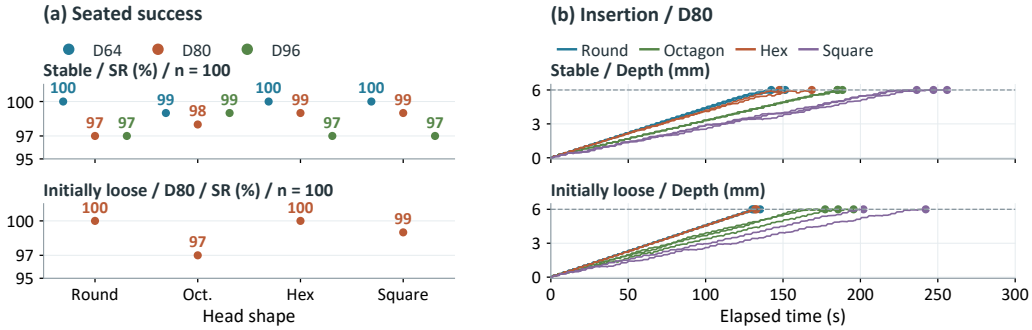

Figure 4: Completion across task conditions and recorded insertion. (a) Seated success over 100 trials per condition; labels show successful trials, with a common denominator of 100. The success-rate axis is zoomed to 95–100%. (b) D80 insertion trajectories ending at their measured endpoints; dashed line: 6 mm target.

5.4 Contact supervision and tactile adaptation

Removing human evidence preserves near-complete progress but increases successful completion time by 11.3 s/5.6 s under stable/loose engagement (Table 1), indicating an efficiency benefit.

Normal-only touch retains similar completion times but increases backdrive to 60.5°/80.2°, supporting directional information for reducing lost motion. Stale feedback reduces PR to 85.5%/73.6% and raises loose backdrive to 114.8°: continuous features alone do not replace current feedback.

Value-only retains long-horizon utility regression but removes both outcome and preference losses (Eq. (10)). PR falls to 81.2%/75.7%, with greater backdrive. This supports their combined supervision, not either loss individually; removing preferences also removes the human-guidance pathway.

Figure 5 illustrates the temporal coupling of load, wrist motion, and insertion over repeated contact renewal.

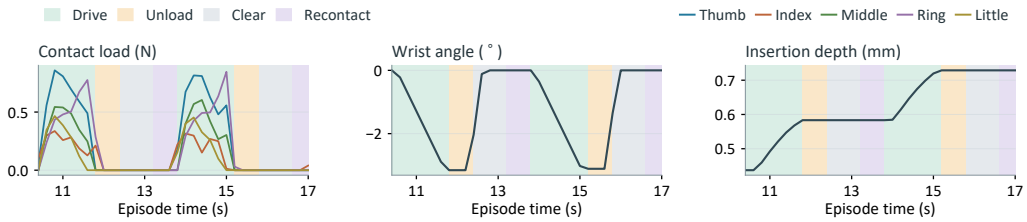

Figure 5: Two recorded round D80 contact cycles: finger-force magnitudes, measured wrist angle, and insertion. Shading denotes controller phase.

5.5 Interpreting long-horizon performance

Completion, time, and backdrive characterize different failure modes. Low backdrive is useful only when accompanied by sustained insertion; an inactive grasp can avoid reversal without completing the task. Likewise, completion time alone favors methods that succeed only on easier conditions because it excludes failures. Reading these measures together, T-FRAME combines high terminal progress with efficient successful execution, rather than obtaining low reversal through inactivity or speed through premature termination. The normal-only ablation further shows that comparable completion times can conceal substantially different amounts of lost motion.

The engagement settings probe complementary contact demands. Stable support isolates repeated traction and renewal around the head, whereas initial play also permits lateral and angular deviations during loading. High completion across both settings therefore concerns more than maintaining an initial grasp. The larger progress loss from stale feedback in loose engagement is consistent with a greater need to adjust commands to the evolving contact state, although the aggregate scores do not localize individual corrections.

The learning result is also more specific than a benefit from longer-horizon scalar supervision: value-only retains the continuation target. Its lower completion indicates that physical-outcome and preference supervision together improve how the available alternatives are used. GPU search supplies those alternatives; the recurrent actor must still realize productive contact transitions using current observations.

6 Limitations and Future Work

Residual failures include contact loss or backdrive under abrupt seating loads, and occasional little-finger entrapment at head corners (one or two runs in some 100-trial batches). More responsive load regulation and clearance-aware recovery remain important for reliable seating.

The reduced model assumes correct thread registration, without explicit flank contacts, cross-threading, or progressive tooth engagement. Video does not identify internal forces, and finite-scenario search gives no general safety guarantee. Physical deployment requires validation beyond simulation. Extending to *grasp–relocate–align–insert–rotate* with a manipulator adds arm–wrist coordination, thread-start search, and misalignment recovery.

7 Conclusion

T-FRAME learns to advance a fastener while preserving the next productive contact. Uncertain human evidence and physically evaluated action alternatives supervise contact outcomes and preferences, which transfer to a recurrent tactile policy without online search. Across simulated head geometries, diameters, and engagement conditions, the method combines near-complete insertion progress with faster successful execution than the adapted baselines. Ablations show complementary benefits from human guidance, directional touch, and current feedback. The central lesson is that sustained fastening depends on learning both what an action achieves and the contact conditions it leaves for the next operation.

REFERENCES

Akshay Agrawal, Brandon Amos, Shane Barratt, Stephen Boyd, Steven Diamond, and J. Zico Kolter. Differentiable convex optimization layers. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/9ce3c52fc54362e22053399d3181c638-Abstract.html.

Xiaoyang Chen, Shengcheng Luo, Haoran Guo, Jiaming Jiang, Wanlin Li, Ziyuan Jiao, and Chenxi Xiao. DexMani: Human-derived manipulability guidance for dexterous rotation. arXiv preprint arXiv:2608.00554, 2026. URL https://arxiv.org/abs/2608.00554.

Ruoxuan Feng, Jiangyu Hu, Wenke Xia, Tianci Gao, Ao Shen, Yuhao Sun, Bin Fang, and Di Hu. AnyTouch: Learning unified static-dynamic representation across multiple visuo-tactile sensors. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/4d893f766ab60e5337659b9e71883af4-Abstract-Conference.html.

Nick Hansen, Hao Su, and Xiaolong Wang. TD-MPC2: Scalable, robust world models for continuous control. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/cf73d57b6dcda32b293df7c2d5341f49-Abstract-Conference.html.

Carolina Higuera, Akash Sharma, Chaithanya Krishna Bodduluri, Taosha Fan, Patrick Lancaster, Mrinal Kalakrishnan, Michael Kaess, Byron Boots, Mike Lambeta, Tingfan Wu, and Mustafa Mukadam. Sparsh: Self-supervised touch representations for vision-based tactile sensing. In *Proceedings of the 8th Conference on Robot Learning*, volume 270, pp. 885–915. PMLR, 2025. URL https://proceedings.mlr.press/v270/higuera25a.html.

Ryan Hoque, Peide Huang, David J. Yoon, Mouli Sivapurapu, and Jian Zhang. EgoDex: Learning dexterous manipulation from large-scale egocentric video. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=FFxkFMU89E.

Elvis Hsieh, Wen-Han Hsieh, Yen-Jen Wang, Toru Lin, Jitendra Malik, Koushil Sreenath, and Haozhi Qi. Learning dexterous manipulation skills from imperfect simulations. In *IEEE International Conference on Robotics and Automation*, 2026. URL https://arxiv.org/abs/2512.02011.

Jialei Huang, Yang Ye, Yuanqing Gong, Xuezhou Zhu, Yang Gao, and Kaifeng Zhang. Spatially anchored tactile awareness for robust dexterous manipulation. arXiv preprint arXiv:2510.14647, 2025. URL https://arxiv.org/abs/2510.14647.

Andela Ilic, Rachel Schuchert, Yijing Jiang, and Christian Holz. EgoPHI: Estimating 3D hand-object contact and force from egocentric vision. In *European Conference on Computer Vision*, 2026. URL https://siplab.org/projects/EgoPHI.

Kushal Kedia, Tyler Ga Wei Lum, Jeannette Bohg, and C. Karen Liu. SimToolReal: An object-centric policy for zero-shot dexterous tool manipulation. arXiv preprint arXiv:2602.16863, 2026. URL https://arxiv.org/abs/2602.16863.

Qingtao Liu, Yu Cui, Zhengnan Sun, Gaofeng Li, Jiming Chen, and Qi Ye. VTDexManip: A dataset and benchmark for visual-tactile pretraining and dexterous manipulation with reinforcement learning. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e19b6f65791e350347bcff8a3955cb5b-Abstract-Conference.html.

Tyler Ga Wei Lum, Kushal Kedia, C. Karen Liu, and Jeannette Bohg. Play2Perfect: What matters in dexterous play pretraining for precise assembly? arXiv preprint arXiv:2606.26428, 2026. URL https://arxiv.org/abs/2606.26428.

Jayanta Mandi, Víctor Bucarey, Maxime Mulamba Ke Tchomba, and Tias Guns. Decision-focused learning: Through the lens of learning to rank. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 14935–14947. PMLR, 2022. URL https://proceedings.mlr.press/v162/mandi22a.html.

Yashraj Narang, Kier Storey, Iretiayo Akinola, Miles Macklin, Philipp Reist, Lukasz Wawrzyniak, Yunrong Guo, Adam Moravanszky, Gavriel State, Michelle Lu, Ankur Handa, and Dieter Fox. Factory: Fast contact for robotic assembly. In *Proceedings of Robotics: Science and Systems*, 2022. doi: 10.15607/RSS.2022.XVIII.035. URL https://www.roboticsproceedings.org/rss18/p035.html.

Jiahe Pan, Stelian Coros, Jitendra Malik, and Toru Lin. Beyond binary: Sim-to-real dexterous manipulation with physics-grounded contact representation. arXiv preprint arXiv:2605.28812, 2026. URL https://arxiv.org/abs/2605.28812.

Haozhi Qi, Brent Yi, Sudharshan Suresh, Mike Lambeta, Yi Ma, Roberto Calandra, and Jitendra Malik. General in-hand object rotation with vision and touch. In *Proceedings of the 7th Conference on Robot Learning*, volume 229, pp. 2549–2564. PMLR, 2023. URL https://proceedings.mlr.press/v229/qi23a.html.

Jie Ren, Zhehao Jiang, Yinhong Yang, Haorui Jia, Han Jiang, Ben Li, Yao Yao, Cheng Lin, Qiu Shen, Zhenshan Bing, Xiao-Xiao Long, and Xun Cao. C2Dex: Contact-consistent reconstruction and retargeting for dexterous manipulation from monocular video. arXiv preprint arXiv:2608.07045, 2026. URL https://arxiv.org/abs/2608.07045.

R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000. doi: 10.21314/JOR.2000.038. URL https://doi.org/10.21314/JOR.2000.038.

Andrei A. Rusu, Sergio Gomez Colmenarejo, Caglar Gulcehre, Guillaume Desjardins, James Kirkpatrick, Razvan Pascanu, Volodymyr Mnih, Koray Kavukcuoglu, and Raia Hadsell. Policy distillation. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1511.06295.

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://proceedings.iclr.cc/paper_files/paper/2024/hash/1cc70be9fb6a83bc46cf4ac21a91e0b0-Abstract-Conference.html.

Balakumar Sundaralingam, Siva Kumar Sastry Hari, Adam Fishman, Caelan Garrett, Karl Van Wyk, Valts Blukis, Alexander Millane, Helen Oleynikova, Ankur Handa, Fabio Ramos, Nathan Ratliff, and Dieter Fox. CuRobo: Parallelized collision-free robot motion generation. In *IEEE International Conference on Robotics and Automation*, pp. 8112–8119, 2023. doi: 10.1109/ICRA48891.2023.10160765. URL https://curobo.org/research/research.html.

Jiawei Wang, Arshiya Taj Abdul, and Evangelos A. Theodorou. cuNRTO: GPU-accelerated nonlinear robust trajectory optimization. arXiv preprint arXiv:2603.02642, 2026. URL https://arxiv.org/abs/2603.02642.

Xinyue Xu, Jieqiang Sun, Jing (Daisy) Dai, Siyuan Chen, Lanjie Ma, Ke Sun, Bin Zhao, Jianbo Yuan, Sheng Yi, Haohua Zhu, and Yiwen Lu. DexCanvas: Bridging human demonstrations and robot learning for dexterous manipulation. arXiv preprint arXiv:2510.15786, 2025. URL https://arxiv.org/abs/2510.15786.

Max Yang, Chenghua Lu, Alex Church, Yijiong Lin, Christopher J. Ford, Haoran Li, Efi Psomopoulou, David A. W. Barton, and Nathan F. Lepora. AnyRotate: Gravity-invariant in-hand object rotation with sim-to-real touch. In *Proceedings of the 8th Conference on Robot Learning*, volume 270, pp. 4727–4747. PMLR, 2025. URL https://proceedings.mlr.press/v270/yang25c.html.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=-2FCwDKRREu.

Wenkang Zhang, Chengbo Yuan, Zicheng Zhang, Zhengxue Cheng, and Yang Gao. EgoTac: In-the-wild tactile prediction from egocentric vision. arXiv preprint arXiv:2608.15060, 2026. URL https://arxiv.org/abs/2608.15060.

Xinghao Zhu, Zixi Liu, Shalin Jain, Chenran Li, Milad Noori, Michael Andres Lin, Huihua Zhao, John Welsh, Mrinal Verghese, Wei Liu, Tingwu Wang, Xingye Da, Zhengyi Luo, Vishal Kulkarni, Naema Bhatti, Yuke Zhu, Linxi Fan, Bowen Wen, Danfei Xu, Soha Pouya, and Yan Chang. Learning dexterous manipulation using contact wrench guidance from human demonstration. arXiv preprint arXiv:2607.00033, 2026. URL https://arxiv.org/abs/2607.00033.

AI USE STATEMENT

Generative AI was used to aid and polish writing, including organizing the presentation, revising prose for clarity and concision, and improving terminological consistency. It also assisted in developing proofs of mathematical claims, including intermediate derivations, examination of assumptions, and checks of the finite-candidate decision analysis. The authors retain full responsibility for the final text and the correctness of the mathematical claims and proofs.

A MECHANICS AND CONTACT INFERENCE

A.1 COORDINATE AND SCALE CONVENTIONS

Mechanics below use world-frame vectors and center-of-mass wrenches ordered as $w = [F; \tau]$, dual to $V = [v; \omega]$. Let p_A be a point on the fixed fixture axis, $b = p_{\mathrm{com}} - p_A$, $r_i = x_i - p_{\mathrm{com}}$, and $\rho_i = x_i - p_A = r_i + b$. Shifting the nominal screw twist from the axis to the center of mass gives

$$t_{\psi,c} = d \begin{bmatrix} (\ell/2\pi)\hat{a} + \hat{a} \times b \\ \hat{a} \end{bmatrix}, \qquad t_{\psi,c}^{\top} w_c = t_{\psi}^{\top} w_A. \tag{14}$$

Here $w_A = [F; \tau_c + b \times F]$. This power-preserving shift matters for offset parts: fixture-axis torque uses ρ_i, whereas the grasp matrix about the center of mass uses r_i. Object-frame features rotate all vectors, including the task axis, by $R_o^{\top}$. Thus $r_i^o = R_o^{\top}(x_i^w - p_{\mathrm{com}}^w)$, $f_i^o = R_o^{\top} f_i^w$, $\rho_i^o = R_o^{\top}(x_i^w - p_A^w)$, and $\hat{a}^o = R_o^{\top}\hat{a}$. The commanded-direction contact torque is $\tau_i^* = d(\hat{a}^o)^{\top}(\rho_i^o \times f_i^o + m_i^{c,o})$, including any modeled patch moment. Compressive normals point into the object, so $f_i = f_i^n n_i + f_i^{\tau}$ has $f_i^n \geq 0$. The physical fields use m, s, N, N m, and rad; normalized learning features and utility terms use positive scales fixed on training data. A contact-force magnitude is not substituted for its normal component.

A.2 FIXED-MODE MECHANICS

For object twist $V = [v; \omega]$, define contact velocity $V_i(V) = v + \omega \times r_i$. Within a fixed-mode segment, contact positions, normals, Jacobians, and a local dynamics linearization are held fixed. For a stable helical segment, $V_k = \alpha_k t_{\psi,c}$ with $\alpha_k \geq 0$ in rad/s. Candidate constraints include

$$\begin{aligned} q_{k+1} &= q_k + \Delta t\, \dot{q}_k, & J_i \dot{q}_k &= V_i(V_k) + \epsilon_{i,k}, \quad i \in \mathcal{C}_{\mathrm{stick}}, \\ n_i^{\top}(J_i \dot{q}_k - V_i(V_k)) &= 0, & & \quad i \in \mathcal{C}_{\mathrm{slide}}, \\ G f_k + B_{\psi}\lambda_{\mathrm{fix},k} + w_g &= M\dot{V}_k + h_{\mathrm{dyn}} + w_{\mathrm{res},k}, & t_{\psi,c}^{\top} B_{\psi} &= 0, \\ f_{i,k}^n &\geq 0, & \|f_{i,k}^{\tau}\|_2 &\leq \mu_i f_{i,k}^n, \\ \tau_k &= J^{\top} f_k + \tau_{\mathrm{dyn},k}, & (q_k, \dot{q}_k, \tau_k) &\in \mathcal{A}_{\mathrm{act}}. \end{aligned} \tag{15}$$

The ideal stable-fixture reaction is reciprocal to the screw motion. Loose support additionally permits transverse and tilt motion, with reactions determined by clearance contacts rather than imposing this ideal reciprocity. The generalized resistance $t_{\psi,c}^{\top} w_{\mathrm{res}}$ is an effort per unit angular progress and contains both running and seating terms. The actual resistive wrench is $-w_{\mathrm{res}}$. With v_i^{τ} defined as hand minus object velocity and force acting on the object, fixed sliding directions impose $f_i^{\tau,o} = \mu_i f_i^n \hat{v}_i^{\tau}$; release imposes zero load and separating velocity. Convex torque–speed envelopes

exclude simultaneous unattainable peak torque and speed. These constraints define a local reference problem, not a global solution across contact creation and destruction.

Initial force states agree with measured load. Normal-force rate bounds and tangent-frame force-increment bounds limit unrealistic instantaneous redistribution,

$$\begin{aligned} -\rho_{i,-}^{n}\Delta t \le f_{i,k+1}^{n} - f_{i,k}^{n} \le \rho_{i,+}^{n}\Delta t, \\ \|f_{i,k+1}^{\tau} - \mathcal{P}_{i,k} f_{i,k}^{\tau}\|_2 \le \rho_i^{\tau}\Delta t. \end{aligned} \tag{16}$$

These convex rate bounds do not replace physical evaluation of mode changes.

A.3 Batched compliant proposals

Parallel proposal generation uses native control trajectories and a compliant surrogate. For reference posture $q^{\mathrm{ref},0}$ and safe joint limits, normalized actions parameterize absolute targets,

$$\begin{aligned} q_k^{\mathrm{goal}} &= q^{\mathrm{ref},0} + [a_k]_+ \odot (\bar{q} - q^{\mathrm{ref},0}) - [-a_k]_+ \odot (q^{\mathrm{ref},0} - \underline{q}), \\ q_{k+1}^{\mathrm{ref}} &= q_k^{\mathrm{ref}} + \mathrm{clip}(q_k^{\mathrm{goal}} - q_k^{\mathrm{ref}}, -\dot{q}_{\max}^{\mathrm{ref}}\Delta t, \dot{q}_{\max}^{\mathrm{ref}}\Delta t). \end{aligned} \tag{17}$$

Actuator response, local compression, and tangential memory predict changes in load. A frozen-object local approximation uses $J_i\dot{q}$ as contact velocity; when object motion is included, the update uses $J_i\dot{q} - V_i(V_o)$. The physical evaluator always advances both hand and fastener. Trajectory, friction, compliance, resistance, and actuation scenarios are batched without treating predicted object motion as a direct control.

The search distribution is defined over bounded temporal knot offsets. At iteration l, samples perturb a warm start u^0:

$$\Delta^{(j)} \sim \mathrm{clip}\left(\mathcal{N}(m_l, \mathrm{diag}(\sigma_l^2)), -\rho_a, \rho_a\right), \qquad u^{(j)} = \mathrm{clip}\left(u^0 + \mathrm{Interp}(\Delta^{(j)}), -1, 1\right). \tag{18}$$

The dimensionless bound ρ_a limits knot offsets. Let $\mathcal{E}_l$ contain the elite offsets under the local objective at iteration l. Its update is

$$\begin{aligned} m_{l+1} &= (1-\alpha)m_l + \alpha\,\mathrm{Mean}(\mathcal{E}_l), \\ \sigma_{l+1} &= \max\{\sigma_{\min}, (1-\alpha)\sigma_l + \alpha\,\mathrm{Std}(\mathcal{E}_l)\}. \end{aligned} \tag{19}$$

Each iteration retains the incumbent. Diagonal sampling, linear knot interpolation, smoothing, and the variance floor regularize the proposal distribution. Random search matches the evaluation count without adapting its distribution. Friction projection acts on predicted shear, independently of action clipping.

The upper-tail CVaR in Eqs. (5) and (9) (Rockafellar & Uryasev, 2000) uses M equally weighted scenarios:

$$\widehat{\mathrm{CVaR}}_\beta(C) = \min_\nu \left[\nu + \frac{1}{(1-\beta)M}\sum_{j=1}^{M}(C_{\zeta_j} - \nu)_+\right]. \tag{20}$$

Replacing the sample average by $\mathbb{E}_\zeta$ gives the population definition. All methods share tail probability and cost scales, including fractional boundary mass when $(1-\beta)M$ is noninteger. Search uses local-model costs; continuation labels use physical execution.

A.4 Task-conditioned manipulation resources

The normalized joint-limit margin in Eq. (6) is

$$r_q = \min_j \frac{\min(q_j - \underline{q}_j, \bar{q}_j - q_j)}{\bar{q}_j - \underline{q}_j}. \tag{21}$$

The motion diagnostic fits contact displacement per unit commanded screw rotation. For a contact at ρ_i relative to the fixture axis, define $y_i = d[(\ell/2\pi)\hat{a} + \hat{a}\times\rho_i]$. Stack the contact Jacobians and target directions, and normalize both by a fixed reference length L_0, $\widetilde{J} = J/L_0$ and $\widetilde{y} = y/L_0$. Then

$$v_q = \widetilde{J}^\top(\widetilde{J}\widetilde{J}^\top + \epsilon_J I)^{-1}\widetilde{y}, \qquad m_\psi = \max_{\alpha\ge 0}\{\alpha : \underline{q} \le q + \alpha v_q \le \bar{q},\ |\alpha v_q| \le \Delta q_{\max}\}. \tag{22}$$

The damping ϵ_J is dimensionless; m_ψ is a step in radians of commanded progress, and v_q maps it to joint displacement. The joint margin r_q in Eq. (6) lies in $[0, 1/2]$ within the limits. We set $m_\psi = 0$ if there is no active contact, no nonzero actuator response, or if the contact-velocity residual exceeds the stated tolerance. This is a local step bound along a damped least-squares direction with sticking contacts, not the global maximum over rolling and sliding contact modes.

For point contacts at measured normal loads, the upper task-axis torque is

$$\tau_{\text{avail}} = \sum_{i \in \mathcal{C}} \left[f_i^n (d\hat{a} \times \rho_i)^\top n_i + \mu_i f_i^n \left\| (I - n_i n_i^\top)(d\hat{a} \times \rho_i) \right\|_2 \right]. \tag{23}$$

This is the support of independent friction disks at fixed normal loads, in N m, not a simultaneous wrench/actuator-feasibility certificate. It measures the rotational component; helical effort also includes axial work. Finite-patch variants replace each point-contact term by the corresponding coupled wrench support; the modeled patch moment is not added twice. With required positive torque $\tau_{\text{req}} > 0$, $r_\tau = \tau_{\text{avail}}/\tau_{\text{req}}$ and

$$r_{\text{release}} = \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \mathbf{1}[\tau_{\text{avail}}^{(-i)} \geq \tau_{\text{req}}], \tag{24}$$

with value zero for an empty contact set. The removal test retains the other contacts' current normal loads. It measures local task-torque reserve, not full force closure or collision-free recontact. The fixture-conditioned model and physical continuation evaluate those additional interaction effects. All resource scales and aggregation weights are fixed on training data. The terminal labels include both r^+ and $r^+ - r$, so the learner can distinguish improving a poor state from consuming an initially large reserve.

A.5 Human transition evidence

EgoDex sequences (Hoque et al., 2026) provide RGB and hand-motion context. Figure 2 uses the finger-operated fixture sequence identified in Table 3; its measured kinematics and illustrative tactile hypotheses are distinct.

Equation (2) uses stiffness k^τ in N/m and damping c^τ in N s/m. With force acting on the object, sliding friction follows the moving hand surface. For an active sliding hypothesis with $k^\tau > 0$, the displacement after projection is

$$\delta_{t+1}^\tau = \frac{f_{t+1}^{\tau,o} - c^\tau \mathcal{P}_t v_t^\tau}{k^\tau}. \tag{25}$$

An unsaturated contact retains $\widetilde{\delta}_{t+1}^\tau$; release resets the history. With zero stiffness, no elastic memory is retained in the saturated update. Sliding, occlusion, and correspondence changes increase uncertainty. Friction and compliance variations are retained when multiple hypotheses explain the observations.

The projection P_Ω compares contact events, surface-motion direction, and pressure trends. Correspondence uses radial/axial position normalized by head radius/height and the task-frame normal, not finger identity. Patch tracks persist only while position and normal matches remain supported; normalization does not equate different collision surfaces.

Window retrieval first matches task direction and the observed prefix within the training video bank. Candidate comparisons then use monotone dynamic programming on the retained channels, with a bounded temporal warp; the same alignment rules apply to every candidate. Event channels use squared probability differences. Observable motion directions use cosine distance; inactive and near-zero-velocity contacts have no direction loss. Pressure changes use a robust distance after within-window centering and scaling, separately for the human relative-pressure signal and robot normal-load trace. Constant traces carry no pressure-change evidence. This preserves temporal structure without equating a visual score to a physical load share.

The weights ω_{kc} combine human-channel validity with available tracking confidence and hypothesis agreement. A missing robot contact is encoded as a contact event, not used to suppress an otherwise valid human event comparison. The human evidence mask and window strength are shared across candidate comparisons. Their reduction to a window-level strength uses a fixed channel/time grid:

$$c_H = \frac{1}{K N_{\text{ch}}} \sum_{k=1}^{K} \sum_{c=1}^{N_{\text{ch}}} \omega_{kc}, \qquad \lambda_H = \lambda_{\max} c_H, \quad 0 \leq \lambda_{\max} < 1. \tag{26}$$

Here N_{ch} counts channels, distinct from angular drawdown D. Missing channels contribute zero, so dropping evidence cannot increase its total weight. The candidate-shared disagreement $\bar{\sigma}_H$ in Eq. (4) averages the variation of agreement scores across retained hypotheses. These are reliability weights, not automatically calibrated probabilities. The human weight is zero when no usable dimension remains, preserving the physical-outcome target. Window selection, comparison scales, and warp settings are fixed before test evaluation.

Geometry and contact timing do not identify internal force. Calibrating force intervals requires synthetic or instrumented targets; integrating pressure requires patch area and force scale. Without them, scale remains a latent hypothesis. Unknown human-side fixture reactions remain uncertain, rather than being set to zero or assigned to hand forces.

A.6 Contact encoding and policy training

Each training candidate includes its native command sequence, duration, and a drive, unload, clear, or recontact intent. Different trajectories may share an intent; candidate identities are local to each state, not fixed online action classes. Refreshed outcome batches retain the identity of the frozen continuation policy that generated their labels. Complete candidate groups stay in one data split to prevent sibling-branch leakage.

Let $b_{i,t}$ denote onset/release indicators and $\Delta t_{i,t}$ contact age. The outcome model uses contact tokens and a recurrent state

$$\begin{aligned} c_{i,t} &= [r_i^o, n_i^o, f_i^n, f_i^\tau, v_i^\tau, \Delta t_{i,t}, b_{i,t}], \\ h_t &= \mathrm{GRU}(h_{t-1}, [\mathrm{Pool}_i\, e_\theta(c_{i,t}), q_t, \dot{q}_t, o_t, \psi])\,. \end{aligned} \tag{27}$$

Pooling is invariant to contact order; fixed training scales preserve geometry and absolute load. Observations o_t contain hand-relative object pose/velocity, insertion, and task axis. Continuous outcome components use factorized regression likelihoods; contact/failure events use classification likelihoods. Missing terminal measurements mask their regression terms without removing the observed failure label. The execution actor maintains its own causal recurrent state, separate from the candidate-conditioned teacher. The action-distillation loss in Eq. (13) is Smooth L1 in normalized executed-command coordinates. A recorded robot motion pattern can initialize the actor or its residual reference without supplying future contact measurements. Teacher parameters remain fixed within an actor update stage, and selected-trajectory supervision regularizes task refinement.

B Finite-Candidate Decision Analysis

The analysis concerns a fixed physical state and a finite candidate set $\mathcal{U}$ of size N. Let $Q(u) = Q_H(u;\pi_b)$ be the fixed continuation utility, and define the effective teacher score

$$S(u) = (1-\lambda_H)Q(u) + \lambda_H \frac{\tau_R}{T_{\mathrm{vid}}} U_H(u). \tag{28}$$

Equation (4) implies $q_{\mathrm{teach}}(u) = \exp(S(u)/\tau_R)/Z$. The score is defined up to an additive constant, which does not affect preferences. All scores are finite on the common candidate support.

Let p be a learned selection distribution, $\Delta_S = \max S - \min S$, and $\epsilon_H = \max(Q-S) - \min(Q-S)$. Gibbs optimality and entropy bounded by $\log N$ give

$$\max_u S(u) - \mathbb{E}_p S(u) \le \tau_R \log N + \Delta_S \sqrt{\tfrac{1}{2}\mathrm{KL}(q_{\mathrm{teach}} \| p)}. \tag{29}$$

To see this, first compare the maximum score with its expectation under the Gibbs teacher; their difference is at most $\tau_R H(q_{\mathrm{teach}})$. The remaining difference in expectations is at most $\Delta_S\, \mathrm{TV}(q_{\mathrm{teach}}, p)$, and Pinsker's inequality gives the second term.

Combining this result with candidate coverage yields

$$V^*_{\mathrm{adm}} - \mathbb{E}_p Q(u) \le \epsilon_{\mathrm{support}} + \epsilon_H + \tau_R \log N + \Delta_S \sqrt{\tfrac{1}{2}\mathrm{KL}(q_{\mathrm{teach}} \| p)}. \tag{30}$$

Indeed, writing $Q = S + (Q - S)$ bounds the additional difference by the range ϵ_H. For $\lambda_H = 0$, the human-score mismatch vanishes. For nonzero λ_H, human preference must justify its contribution: teacher imitation alone does not guarantee improvement in robot utility.

The selection distribution is the candidate preference model π_θ; the continuous execution actor is not a categorical distribution over this set. With numerical probability smoothing, use $S = \tau_R \log q_{\text{teach}}$ and recompute ϵ_H. The result bounds expected selection rather than greedy argmax performance. It does not bound action-distillation error, errors in physical outcome estimation, or long-horizon task success. Its role is to separate candidate coverage, teacher alignment, and selection error rather than attributing all downstream differences to the representation alone.

C Fastener Model and Engagement Conditions

The moving head and threaded shank share one rigid-body pose and combined mass properties. For components with common-frame inertia I_b, mass m_b and center c_b,

$$c = \frac{\sum_b m_b c_b}{\sum_b m_b}, \qquad I_c = \sum_b \left[I_b + m_b \big(\|c_b - c\|^2 I - (c_b - c)(c_b - c)^\top \big) \right]. \tag{31}$$

Component tensors are rotated into the common frame before summation. Gravity acts at this center of mass; an off-axis or tilted head can therefore change support loading. External head geometry determines fingertip contacts, while the screw coordinate determines nominal axial feed.

Stable support maintains axis alignment. The initially loose reference uses two unilateral support rings separated by $L_s = 4\,\text{mm}$, with radial clearance $c_r = 18.75\,\mu\text{m}$. Each ring uses eight transverse directions. Translation and tilt consume the same clearance budget; the head is not snapped back onto the axis during motion. A centered small-angle estimate is

$$\beta_{\text{free}} \simeq 2c_r / L_s \simeq 0.537^\circ. \tag{32}$$

This value describes available play, not a measured tilt trajectory or a fraction of thread engagement. The support span remains fixed through insertion; it does not represent a growing number of loaded thread teeth.

Thread resistance, support reaction, and final shoulder loading are distinct. Running resistance opposes motion, and seating load appears as the shoulder gap closes. Frictional work accumulates with motion; friction force does not accumulate simply because another revolution has occurred. The target-depth seat additionally requires the torque and low-speed hold criterion given in the task configuration.

This reduced model studies repeated contact renewal after correct thread registration. Explicit flank geometry, lead-in search, cross-threading, and deformation require a different fidelity level, such as contact-rich assembly simulation (Narang et al., 2022). They are not inferred from rendered threads or from exact satisfaction of the imposed helical coordinate.

D Experimental Details

D.1 Task distributions and implementation

Table 2 separates grasp geometry, thread kinematics, and support mechanics. The evaluated matrix varies head geometry, grasp diameter, and engagement while fixing nominal thread mechanics. Rotation signs follow Eq. (1).

Model and training conventions follow Appendix A.6. Normalization, resource scales, and utility weights are fixed on training data before held-out evaluation.

D.2 Baseline implementation and comparison scope

Controlled ablations match tactile/task interfaces, history, action permissions, and rewards except for the explicitly removed information or supervision. External-policy comparisons retain method-specific inputs.

External-policy comparisons. Play2Perfect starts from its released screwing policy and fine-tunes on the common task (Lum et al., 2026). SimToolReal tracks object-pose goals derived from the axis,

Table 2: Task factors for the reported execution matrix. Head envelope is not nominal thread diameter; lead includes handedness and start count.

Factor	Definition and evaluated values
Head geometry	Round, octagonal, hexagonal, square
Grasp envelope	D64, D80, D96 (mm); shape-specific collision surface
Thread kinematics	Pitch $p = 2$ mm, $n_s = 1$, $s = -1$; signed lead $\ell = -2$ mm/revolution; mean thread diameter 16 mm
Target	6 mm insertion (three revolutions), seating hold, 450 s deadline
Support	Stable alignment or initial finite lateral/angular play
Contact mechanics	Material friction: object 0.5, pad 0.8, metal 0.1 (average combination); solid-head density 2200 kg/m^3
Thread resistance	Static/kinetic friction 0.15/0.10, axial load 5 N; breakaway/running prevailing torque 0.003/0.002 N m; seating resistance 0.004 N m from 80% insertion
Coverage	Twelve stable shape–diameter cells and four initially loose D80 shapes; fixed lead and thread mechanics, not held-out mechanical splits

Table 3: Simulation and visualization settings. Task values are in Table 2.

Setting	Value
Execution simulation	Isaac Sim/PhysX; physics 240 Hz (stable), 960 Hz (initially loose); environment/tactile sampling 60 Hz, policy 20 Hz; two-step control delay
Video illustration	EgoDex finger-operated fixture, clip 250; 234 frames at 30 fps, dataset hand-pose annotations; tactile fields are schematic, not inferred outputs
Execution statistics	Task-weighted means in Table 1; 100 trials per T-FRAME condition, with success counts in Figure 4

lead, and target depth (Kedia et al., 2026). Goal advancement is fixed before evaluation. Both preserve the released 29-dimensional arm–hand action semantics; method-specific sensing and pretraining are distinguished from downstream adaptation. DexScrew provides the fastening-specific transfer comparison, using its learned rotational skill and tactile-history approach (Hsieh et al., 2026). Table 1 reports adaptation to the target task distribution, rather than native-task scores from the original publications.

Matched-candidate comparisons. To isolate the representation from action generation, matched-candidate evaluation holds the candidate support fixed. Raw touch, CoP/force features (Pan et al., 2026), object net wrench, DexMani-style manipulability, and CHORD-style wrench capability (Zhu et al., 2026) define feature-level comparison models, with candidate encoder, history access, capacity, and training examples held constant. An oracle selects by realized physical continuation value; it is an upper bound within that candidate set, not a globally optimal controller. External-policy adapters retain native arm-increment and absolute hand-target actions, while replacing objects, fixtures, goals, and task rewards. Released weights initialize adaptation.

D.3 Full-task metrics and statistics

Each evaluation starts from a feasible physical state sampled from the common task distribution. An episode succeeds after attaining complete progress, reaching target insertion, and maintaining the specified low-speed seating condition for the prescribed duration. Timeouts and physical failures are failures. For the successful episodes $\mathcal{S}$ within a task cell, completion time is

$$T = \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} T_{\mathrm{seat},i}, \qquad |\mathcal{S}| > 0. \tag{33}$$

It is undefined when $\mathcal{S}$ is empty. Across tasks, time averages weight each task with a defined completion time equally; tasks without successes remain in progress and success statistics. Our evaluation deadline is 450 s. Completion progress uses the terminal valid state, rather than the

Table 4: Supervision in the learning controls. A checkmark retains the corresponding full-method loss; a dash sets its coefficient to zero. No row changes the available tactile observations or candidate data.

Variant	Utility regression	Contact outcomes	Preference CE
T-FRAME	✓	✓	✓
Without contact-outcome supervision	✓	–	✓
Without preference supervision	✓	✓	–
Value-only	✓	–	–

maximum visited angle or accumulated positive motion:

$$\mathrm{PR} = 100\ \mathrm{clip}_{[0,1]}\left(\frac{\xi_T - \xi_0}{\Theta_{\mathrm{req}}}\right). \tag{34}$$

Here ξ is task-directed unwrapped rotation, ξ_0 is its evaluation origin, and $\Theta_{\mathrm{req}} = 1080^\circ$ for the current three-turn suite. Early physical failures retain their last valid progress; PR does not imply successful seating. Measured depth is retained independently, since preload offsets or compliance can make it differ from ideal helical feed. Progress, success, and T are interpreted jointly. Backdrive is $B = \sum_t [-(\xi_{t+1} - \xi_t)]_+$ on a common sampling grid, excluding resets. Maximum drawdown D is a separate candidate-scoring quantity and is not substituted for B in the task tables. Overload duration, slip distance, and recontact recovery supplement the primary metrics.

Each reported T-FRAME task condition contains 100 evaluation trials. Means first average trials within a task cell, then weight cells equally within each engagement regime. Success rates use successful/total trial counts. Evaluation seeds vary trial conditions, not independently trained models.

D.4 CLOSED-LOOP AND HUMAN-EVIDENCE ABLATIONS

Stale-touch evaluation keeps the policy and action interface fixed while replacing current tactile channels by their delayed or frozen counterparts. The physical contact solver remains unchanged. Normal-only inputs remove shear and slip information without removing normal load or position. Paired contact, command, and progress traces distinguish a change in the action from useful recovery.

Loss-removal controls retain the same scoring interface, observations, and candidate groups. Table 4 specifies their supervision; retained terms keep their full-method coefficients and normalization.

Removing human evidence sets $\lambda_H = 0$ while retaining motion warm starts, robot trajectories, and training budgets. Human evidence and all normalization parameters are frozen before policy evaluation.

D.5 CANDIDATE EVALUATION AND COMPUTATION

Each parent and its complete candidate set stay in one data split (Appendix A.6). State cloning includes controller and contact history. The physical evaluator executes each candidate and its common continuation without allowing faster branches to evolve while others finish. Selected failures remain outcomes. Coverage records the best realized value and fraction of states containing a productive feasible action.

For parent s with realized continuation values $Q_s(u)$, regret is

$$\mathcal{R}_s = \max_{u \in \mathcal{U}_s} Q_s(u) - Q_s(\hat{u}_s). \tag{35}$$

Methods use the same candidate support, independent physical evaluation scenarios, and continuation definition. Selection does not use held-out continuation outcomes. The empirical best candidate is a finite-set reference, not an estimate of the global optimum over all trajectories. Regret uses the dimensionless utility in Eq. (9); physical failure remains part of the outcome. Unobserved continuation outcomes are excluded rather than assigned zero regret.

Policy/random proposals, nominal local optimization, and scenario-based optimization define proposal-quality controls at equal physical-rollout budgets. The relevant quantities are best physically

realized continuation value, selected value, and successful recontact, rather than the local surrogate objective alone. Training with and without optimization data retains the same deployed policy interface; neither uses online search.

Feasible-candidate yield is the fraction of physically evaluated candidates meeting the same physical constraints, not merely their surrogate predictions; states with no feasible candidate remain in the denominator of state coverage. Different candidate sets are compared by realized utility and coverage, not regret alone: a uniformly poor candidate set can have small selection regret. GPU batching evaluates action and mechanics hypotheses concurrently during training; the deployed actor uses tactile feedback without trajectory search.

D.6 Video evidence and observability

Contact timing and relative-pressure dynamics are distinct from absolute force calibration. Observable sliding constrains tangential direction, while absolute force validation requires synthetic or instrumented targets. EgoPHI (Ilic et al., 2026) and EgoTac (Zhang et al., 2026) address perception rather than the end-to-end robot-control comparison. Figure 2 illustrates the representation using recorded kinematics and schematic tactile hypotheses; the latter are not measured forces or inferred outputs for that clip.

E Qualitative Contact Sequences

Figures 6–9 show one reference episode per condition, with prescribed recontact timing and condition-specific settings. They illustrate contact renewal, not aggregate policy performance or paired feedback effects. Frames are chronological and nonuniformly spaced; labels report measured cumulative rotation in °.



Figure 6: D64 stable-support reference trajectories. Each row shows contact renewal and subsequent insertion for a different head shape.



Figure 7: D80 stable-support reference trajectories across four head shapes.



Figure 8: D80 initially loose reference trajectories. All four displayed episodes reach the 6 mm target and the reduced-model seating condition. Poses and rotation labels follow the recorded trajectory, without magnifying tilt.



Figure 9: D96 stable-support reference trajectories. All four displayed episodes reach the 6 mm target and seating, including the hexagonal execution with preload-aware recontact. Rendered threads do not imply explicit flank collisions.