RISK-AWARE BENCHMARKING AND ADAPTATION FOR GENTLE RETRIEVAL THROUGH YIELDING OCCLUSIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Retrieving objects through compliant occlusions requires a robot to make contact while limiting disturbance to both the environment and itself. Existing manipulation benchmarks largely emphasize task completion or aggregate cost, making it difficult to diagnose failures caused by changing contact mechanics, retrieval function, and brief high-risk events. We introduce GRAM, a benchmark for gentle retrieval through yielding occlusions with Expose–Attach–Transport diagnostics, controlled morphology shifts in thin-shell and branching occluders, and explicit environment-disturbance and robot self-preservation budgets. GRAM reveals that mechanics-relevant shifts sharply degrade gentle retrieval, failures localize differently across retrieval stages and risk channels, and budget violations are often driven by short peak events. From these findings, we derive SRAH, a lightweight residual adapter that conditions a frozen policy on retrieval stage, recent mechanics response, remaining budget headroom, and short-horizon hazard, while using unsafe and failed rollouts as counterexample supervision. Across ACT-style, diffusion, and VLA backbones, SRAH improves gentle success by 4.1–7.1 percentage points under cross-family and tightened-budget evaluation while reducing environment- and robot-side violations. Targeted physical analogues show the same qualitative pattern, with larger improvements in acceptable retrieval than in raw completion.

1 INTRODUCTION

Robotic retrieval often requires contact with the structure that hides a target. A robot may need to open a fold, separate compliant packaging, or move through foliage before it can grasp and extract the desired object. In these settings, the occluding structure is neither rigid clutter to be avoided nor the object whose final configuration defines the task; it is a yielding medium that must be reconfigured only enough to make retrieval possible. This creates a coupled requirement: the robot must make progress through contact while limiting collateral disturbance to the environment and excessive loads on itself. We study this regime as *retrieval through yielding occlusions*.

Modern visuomotor policies provide strong manipulation backbones, and deformable-manipulation benchmarks provide increasingly realistic contact-rich environments (Zhao et al., 2023; Chi et al., 2025; Kim et al., 2025; Lin et al., 2020; Lu et al., 2024). However, existing evaluations usually emphasize task completion, generic scene variation, or manipulation of the deformable object itself (James et al., 2020; Gu et al., 2023; Liu et al., 2023a). They therefore provide limited leverage for answering three questions that become central when the deformable structure is an intervening occluder: where in the retrieval process a policy fails, whether failure is caused by a change in interaction mechanics rather than appearance, and whether a nominally successful trajectory violates a disturbance or self-preservation constraint. Constrained learning methods can reduce aggregate costs (Lee et al., 2022), but an aggregate budget does not by itself identify which contact behavior should change or when.

We first introduce **GRAM**, a benchmark designed to make these failure factors measurable. GRAM factorizes retrieval into EXPOSE, ATTACH, and TRANSPORT functions; generates two complementary families of yielding occlusions with controlled morphology and mechanics variation; and defines success as target retrieval subject to explicit environment-disturbance and robot self-preservation

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.

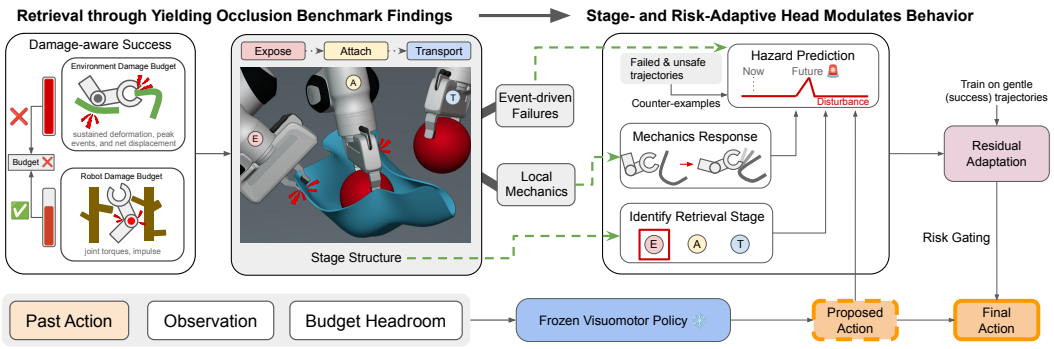

Figure 1: **GRAM reveals the interaction state needed for gentle retrieval, and SRAH operationalizes it.** Controlled mechanics shifts and stage- and budget-resolved diagnostics expose three recurring factors behind failure: retrieval stage, local mechanics response, and imminent peak-event risk. These findings motivate SRAH's stage estimator, interaction-history representation, and short-horizon hazard predictor, which condition a bounded residual correction to an existing policy. Unsafe and failed rollouts supervise the predictive interface without being used as actions to imitate.

budgets. This structure lets us vary mechanics-relevant morphology while holding the robot interface and task semantics fixed, then attribute failure to a retrieval function and risk channel rather than only to end-to-end success. Across learned and control baselines, three patterns recur. First, morphology shifts can substantially change which interaction strategies work. Second, failure modes localize differently across retrieval stages and occluder mechanics. Third, under tight budgets, failure is often caused by short peak events rather than gradual cost accumulation. Together, these results suggest a compact organizing principle: *low-disturbance retrieval depends on knowing what retrieval function is currently being performed, how the local medium is responding to recent contact, and whether the proposed action is likely to trigger an imminent budget-critical event.*

We use this diagnosis to design **SRAH**, a Stage- and Risk-Adaptive Head that modulates the action of a frozen base policy. SRAH predicts the current retrieval stage, summarizes recent interaction history with a mechanics-response latent, tracks remaining environment and robot budget headroom, and predicts short-horizon hazard conditioned on the base action. These signals condition a bounded residual correction rather than replacing the underlying policy. The same decomposition also determines how data are used: gentle successful rollouts provide residual action targets, while unsafe and failed rollouts supervise stage, interaction-response, and hazard prediction without being imitated. SRAH directly test whether the structure exposed by GRAM is useful for policy adaptation.

Our experiments evaluate both sides of this claim. GRAM diagnoses failure under controlled morphology shift, cross-family transfer, and tightened budgets; SRAH is then tested on the regimes and failure modes implicated by those diagnostics. Across diffusion policy, action-chunking and vision-language-action backbones, SRAH improves budgeted gentle success while reducing environment and robot-side violations. Component ablations show that stage, mechanics history, hazard prediction, and counterexample supervision account for distinct parts of the improvement. Targeted physical analogues reproduce the same qualitative pattern: SRAH primarily converts unsafe or borderline completions into acceptable retrievals rather than simply increasing raw completion.

Our contributions are:

- **A diagnostic benchmark for yielding-occlusion retrieval.** GRAM combines stage-resolved retrieval, controlled mechanics-driven morphology shifts, and explicit environment-disturbance and robot self-preservation budgets to expose failure modes hidden by aggregate task success.
- **An empirical characterization of the interaction state that governs failure.** GRAM shows that retrieval failures are structured by functional stage, local mechanics response, and short-horizon peak-event risk, providing a compact explanation for failures under morphology and budget shift.

- **A benchmark-derived adaptive policy.** SRAH turns these findings into a lightweight residual adapter for learned manipulation policies and uses unsafe and failed rollouts as predictive counterexamples rather than demonstrations.

2 Related Work

Manipulation benchmarks and generalization. Manipulation benchmarks standardize evaluation across tasks and environments (James et al., 2020; Yu et al., 2020; Mu et al., 2021; Gu et al., 2023; Kumar et al., 2023; Liu et al., 2023a), while large-scale robot datasets study transfer across scenes, tasks, and embodiments (Walke et al., 2023; O'Neill et al., 2024; Khazatsky et al., 2024). Perturbation and generalization suites broaden this coverage (Szot et al., 2021; Li et al., 2024; Pumacay et al., 2024; Haresh et al., 2024). GRAM complements these efforts by varying mechanics-relevant morphology in a yielding occluder while holding the retrieval objective and control interface fixed, with failures resolved by retrieval stage and risk channel.

Deformable and contact-rich manipulation. Research on cloth, cables, bags, and garments uses learned policies, dynamics models, simulation, and contact feedback (Lin et al., 2020; Huang et al., 2021; Lu et al., 2024; Seita et al., 2021; Wu et al., 2023; Shen et al., 2024; Sunil et al., 2022). Here, the deformable structure is an intervening occluder rather than the object whose final configuration defines success. Reconfiguration matters only insofar as it enables retrieval without excessive disturbance, motivating GRAM's functional diagnostics and separate environment and robot budgets.

Adaptive and risk-aware policies. Spatial, action-chunking, diffusion, and vision-language-action policies provide manipulation backbones (Zeng et al., 2021; Shridhar et al., 2023; Zhao et al., 2023; Chi et al., 2025; Kim et al., 2025); constrained learning incorporates safety costs (Achiam et al., 2017; Lee et al., 2022; Liu et al., 2023b); and context-based adaptation uses interaction history (Kumar et al., 2021; Rakelly et al., 2019). SRAH adapts a frozen backbone using the interaction state implicated by GRAM: retrieval stage, local mechanics response, budget headroom, and short-horizon hazard. Unsafe and failed trajectories supervise these predictions without providing actions to imitate.

3 GRAM: Diagnosing Retrieval Through Yielding Occlusions

GRAM asks which aspects of contact determine whether retrieval through a yielding occluder succeeds gently. It separates task progress from harmful shortcuts, controls mechanics-relevant morphology, and attributes failures to retrieval function and risk channel.

3.1 Defining Acceptable Retrieval

Successful extraction alone cn reward destructive shortcuts. For rollout τ, let $G(\tau)$ indicate that the target reaches and remains in the goal region. GRAM measures environment disturbance D_{env} and robot self-preservation cost C_{robot} relative to budgets B_{env} and B_{robot}, giving $\bar{D}_{\text{env}} = D_{\text{env}}/B_{\text{env}}$ and $\bar{C}_{\text{robot}} = C_{\text{robot}}/B_{\text{robot}}$. A *gentle success* requires

$$\text{Succ}(\tau) = \mathbb{1}\left[G(\tau) \wedge \bar{D}_{\text{env}}(\tau) \leq 1 \wedge \bar{C}_{\text{robot}}(\tau) \leq 1\right].$$

Environment cost combines sustained deformation, localized peaks, and accumulated structural displacement; robot cost combines sustained load, abrupt load changes, and contact impulses. Component normalization and calibration are fixed before evaluation. These quantities are interaction-risk proxies, not material-failure models (Appendix C).

Gentleness must correspond to undesirable interaction, not merely a chosen cost. In blinded human-label evaluation, EnvDist predicts damage-like events with ROC–AUC 0.87 for thin shells and 0.84 for branches, versus 0.74 and 0.70 for displacement alone (Table 8). This supports peak-sensitive disturbance measurement without treating the proxy as a material-failure model.

To identify where behavior fails, GRAM labels three retrieval functions: Expose creates visual and physical access; Attach establishes stable acquisition; Transport extracts and carries the acquired target. Stable attachment takes precedence over access, so renewed occlusion during extraction remains Transport. This Expose–Attach–Transport (EAT) factorization is diagnostic, not a scripted sequence: policies may re-expose, retry attachment, or regrasp (Appendix B.1).

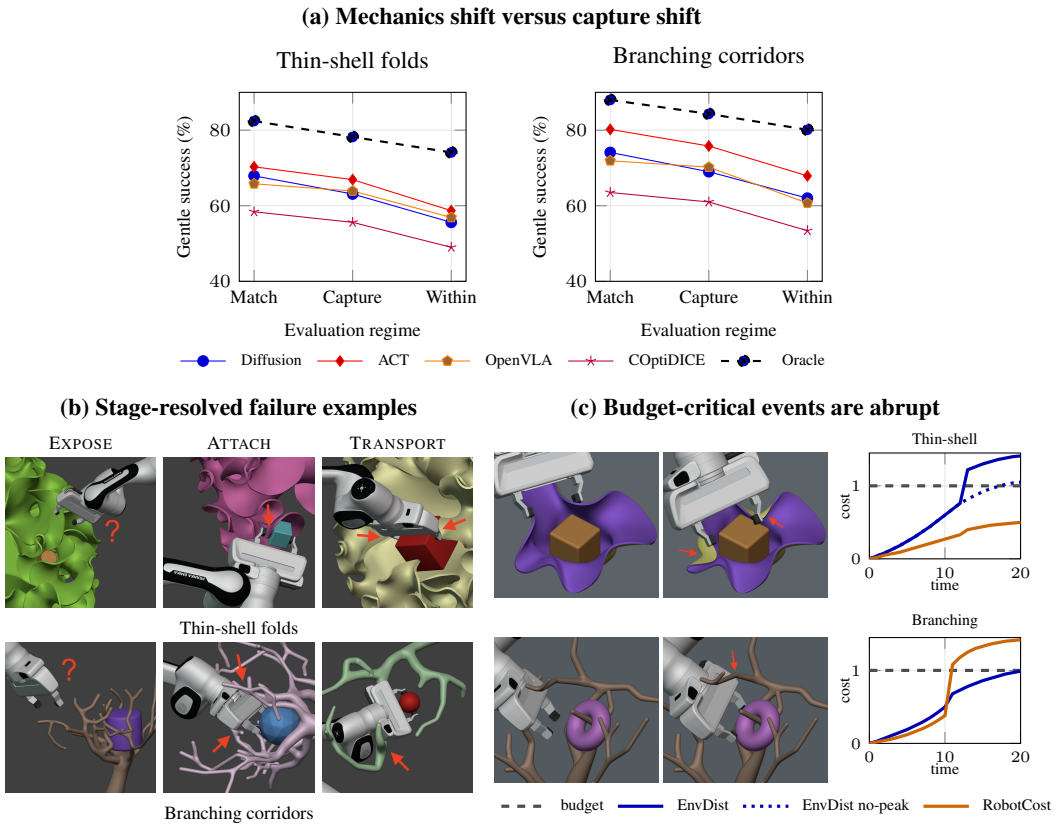

Figure 2: **GRAM diagnoses mechanics-, stage-, and event-dependent failures.** **(a)** Matched, Capture, and Within-family means for the full composite suite (Table 15); Within-family shift reduces gentle success more than Capture shift, including for the privileged oracle. **(b)** EAT distinguishes access, acquisition, and extraction failures. **(c)** Contact events produce abrupt budget crossings. Omitting the thin-shell peak term from the same rollout delays, but does not eliminate, its crosssing.

3.2 CONTROLLED MECHANICS SHIFT

To vary contact challenges without changing the task, two procedural families share the robot, executed action interface, retrieval objective, and budget semantics. **Thin-shell folds** (CABBAGE) vary fold severity, compliance, and stiffness heterogeneity, creating pockets, narrow fold mouths, and localized deformation. **Branching corridors** (BIFURCATOR) vary branch density, tortuosity, and radius, changing clearance and snag-prone extraction paths. These families expose complementary contact challenges: creating access and extracting through continued contact (Appendix B.2).

3.3 DIAGNOSTIC EVALUATION PROTOCOL

We separate changes in mechanics from changes in risk tolerance and visual input. **Matched** tests morphologies from the training support. **Within-family** moves generator parameters toward held-out structural regimes. **Cross-family** trains on one family and evaluates on the other; table columns name the evaluation family. Benchmark fitting of all policy and adapter components uses only the source family. **Tightened-budget** multiplies both budgets by 0.8, without retraining or changing cost calibration; budget-conditioned inputs use the reduced limits. **Capture** perturbs cameras, lighting, and backgrounds while keeping mechanics fixed. Appendix B.3 specifies the distributions.

Each main regime uses 2,000 fixed episodes per family, shared across methods, with stratified-bootstrap confidence intervals. We record raw and gentle success, continuous costs, stage-labeled budget events, and a separate terminal reason.

3.4 Findings That Motivate Adaptation

Mechanics and local response. For Diffusion Policy, ACT, and OpenVLA-ft, Within-family shift reduces gentle success by 8.9–12.3 percentage points, versus 1.7–5.1 under Capture (Fig. 2a). Local aperture and snag descriptors also outperform coarse global geometry as failure predictors (Table 18). These results motivate representing local interaction response rather than relying on scene identity.

Retrieval function. Stage attribution distinguishes thin-shell access failures from branching extraction and snag failures (Fig. 2b). A contact behavior useful for opening access can be inappropriate during extraction; adaptation therefore needs the current retrieval function.

Imminent risk. Fold-mouth and snag events cause abrupt budget crossings (Fig. 2c). Recomputing the thin-shell trace without its peak term delays the crossing but does not prevent it. Headroom measures remaining margin; anticipating its abrupt consumption requires short-horizon risk prediction.

Together, these findings motivate SRAH's *functional* stage, history-based *physical* response, and *temporal* hazard estimates. Budget headroom supplies the constraint context; unsafe and failed traces supervise prediction without supplying actions to imitate. Keeping the backbone fixed tests whether this diagnostic structure improves retrieval.

4 SRAH: Stage- and Risk-Adaptive Retrieval

SRAH tests whether GRAM's diagnostic structure is useful for adaptation: it supplies a frozen policy with functional, physical, and temporal interaction state rather than replacing its task policy (Fig. 1). Let h_t contain current sensing and the preceding eight control steps of executed-action, proprioceptive, and contact history. Freezing the backbone preserves its learned task policy and native action generation while testing an interaction-state adapter across policy families. The central design choice is the information supplied for correction, not a new backbone architecture. Residual adaptation provides a common interface for applying this information across backbone families. It is not a comparison against end-to-end retraining.

The backbone's native inference wrapper supplies its next Cartesian command a_t^0; SRAH executes

$$a_t = \Pi_{\mathcal{A}}(a_t^0 + \lambda_t \Delta a_t), \qquad \lambda_t \in [0, 1], \tag{1}$$

where $\Pi_{\mathcal{A}}$ clips the command to the valid action range.

4.1 Functional Ambiguity: Retrieval Stage

Contact that creates access can obstruct extraction after acquisition. The stage posterior $\hat{z}_t = q_\phi(h_t)$ distinguishes Expose, Attach, and Transport, using GRAM's access/attachment labels for supervision (Appendix B.1). It's condition the residual without prescribing a stage sequence, so re-exposure and regrasping remain possible.

4.2 Physical Ambiguity: Mechanics Response

Similar-looking contact states can respond differently to a probe or pull. Rather than using family identity, a history encoder $m_t = e_\eta(h_t)$ summarizes local interaction response. A consequence-prediction head supervises this latent over 12 control steps: changes in access, target progress, and both risks use regression; attachment stability and snag occurrence use binary classification. These targets favor behaviorally relevant response over static scene identity.

4.3 Temporal Ambiguity: Budget Headroom and Hazard

Remaining margin does not reveal whether the next contact will consume it abruptly. SRAH combines current headroom, $b_t = [1 - \bar{D}_{\mathrm{env},t},\ 1 - \bar{C}_{\mathrm{robot},t}]_{[0,1]}$, with two action-conditioned hazard scores, $\hat{\rho}_t = r_\psi(h_t, a_t^0, \hat{z}_t, m_t, b_t)$. Each predicts a first crossing of its channel's budget within 12 control steps, before that channel has crossed. Both headroom and crossing labels use the active budget, including the reduced Tightened-budget limits. This separates accumulated risk from impending risk.

4.4 Residual Adaptation

The residual $\Delta a_t = g_\theta(h_t, a_t^0, \hat{z}_t, m_t, b_t, \hat{\rho}_t)$ uses these signals to adjust the proposed action. An affine-sigmoid gate combines minimum headroom and maximum hazard; its initialization favors smaller interventions near depleted budgets or predicted crossings. The gate limits the residual, not the backbone action, and provides no safety guarantee. Adaptation preserves each backbone's native replanning and chunk-execution schedule, with executed actions fed back into history (Appendix D.6).

4.5 Learning from Successful and Failed Interaction

Unsafe completions and failed attempts reveal precursors that gentle trajectories miss. Stage, consequence, and hazard losses therefore use all three data partitions, but residual imitation uses only gentle successes, with a deomstrated action $a_t^\star$:

$$\mathcal{L}_{\text{res}} = \mathbb{E}_{\mathcal{D}_{\text{gentle}}} \left[\left\| a_t^\star - (a_t^0 + \lambda_t \Delta a_t) \right\|_2^2 \right]. \tag{2}$$

The gate and residual head are optimized jointly, with penalties on residual magnitude and temporal variation; the backbone is remains frozen. Predictive losses condition on the logged action that generated their targets, whereas deployment queries hazard with the base proposal. Predictions describe logged continuation behavior, not counterfactual SRAH rollouts. Thus failures teach what to anticipate without teaching which actions to repeat. Appendix D details full objective and prediction masks.

5 Results

We test whether the interaction state diagnosed in Section 3 improves gentle retrieval, which components contribute, and where behavior changes. We pair SRAH with Diffusion Policy (Chi et al., 2025), ACT-style (Zhao et al., 2023), and OpenVLA-ft (Kim et al., 2025), keeping each backbone fixed; COptiDICE (Lee et al., 2022) provides a constrained-learning reference. The shared action interface does not require identical internal policy representations (Appendix G.1).

5.1 SRAH Improves Retrieval Under Diagnosed Shifts

Under Cross-family and Tightened-budget evaluation, SRAH improves every backbone in both families by 4.1–7.1 percentage points, averaging 5.2 points, while reducing environment and robot violations under tightened budgets (Table 1). Comparison with COptiDICE depends on the backbone and regime: Diffusion+SRAH has higher Cross-family gentle success than COptiDICE, but lower Tightened-budget success. The consistent result is improvement of each frozen backbone, not uniform superiority over constrained learning.

5.2 Components Address Complementary Needs

Table 2 separates the adapter's inputs and supervision. The first increment adds residual adaptation and stage conditioning jointly; adding budget headroom further reduces violations. Adding hazard to stage+budget improves mean Tightened-budget success by 1.1 points versus 0.1 under Cross-family transfer; adding mechanics to that variant gives 1.9 points under Cross-family transfer versus 0.9 under tightened budgets. This pattern supports temporal anticipation under budget pressure and local response modeling under transfer, without implying improvement in every cell.

Table 3 compares mechanics response with a static family ID. Mechanics gives larger gains in both regimes; the unseen destination ID makes Cross-family a limited test of categorical identity. Gentle-only predictor training reduces Full SRAH's mean gain from 5.2 to 4.0 points while leaving action supervision unchanged. Together, these controls support response-based context and risk-bearing predictive data, rather than family identity or gentle demonstrations alone.

For the budgets defining the data partitions, gentle trajectories contain no budget-crossing positives. The gentle-only control therefore removes predictive event coverage as well as training examples. Its result supports learning from adverse outcomes, but does not separate coverage from dataset size.

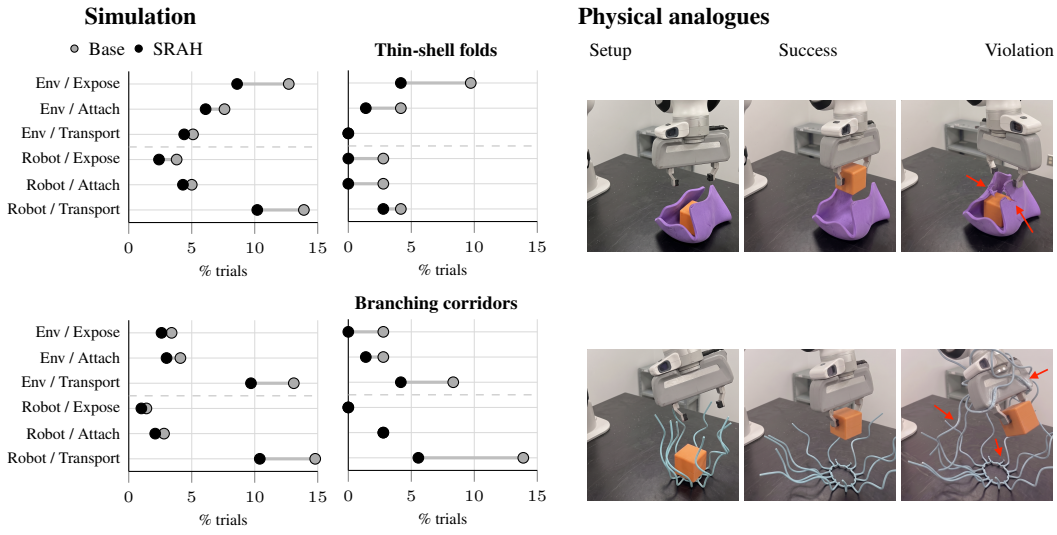

Figure 3: **Stage- and risk-localized changes under SRAH.** Points compare marginal event rates for each risk-channel/stage pair. Thin-shell simulation shows reductions during exposure and transport; branching reductions concentrate in transport. Physical examples show the corresponding access and extraction interactions.

Table 1: **SRAH improves gentle retrieval under mechanics and budget shift.** Simulation reports gentle success (%) with 95% stratified-bootstrap confidence intervals; E/R denotes Tightened-budget environment/robot violation rates. Cross-family columns name the evaluation family. Physical entries are counts out of 24 trials per backbone and analogue.

	Thin-shell folds			Branching corridors		
Method	Cross-family ↑	Tightened ↑	E/R viol. ↓	Cross-family ↑	Tightened ↑	E/R viol. ↓
Diffusion Chi et al. (2025)	31.7±1.8	27.9±1.7	36.0/32.0	35.4±2.0	30.7±1.9	32.0/30.0
+ SRAH	**37.8 ± 1.9**	**35.0 ± 1.8**	**27.8 / 24.6**	**41.5 ± 1.9**	**37.4 ± 1.9**	**24.7 / 22.1**
ACT-style Zhao et al. (2023)	35.9±1.9	45.1±2.0	22.0/20.0	44.9±2.0	49.8±2.0	18.0/16.0
+ SRAH	**41.2±1.8**	**50.8±1.9**	**16.2/14.5**	**49.4±1.9**	**54.4±1.9**	**12.2/10.7**
OpenVLA-ft Kim et al. (2025)	43.8±2.0	52.3±2.0	18.0/16.0	52.0±2.0	54.3±2.0	12.0/11.0
+ SRAH	**48.0±1.9**	**56.4±1.9**	**13.2/11.5**	**56.4±1.9**	**58.4±1.9**	**9.0/7.7**
COptiDICE	33.2±1.9	44.0±2.0	12.0/11.0	36.6±2.0	47.2±2.0	10.0/9.0

		Thin-shell physical analogue		Branching physical analogue	
Backbone	Variant	Raw/Gentle ↑	E/R viol. ↓	Raw/Gentle ↑	E/R viol. ↓
Diffusion Chi et al. (2025)	Base	13/9	4/3	13/7	4/5
	Full SRAH	**14/12**	**2/1**	**14/11**	**2/3**
ACT-style Zhao et al. (2023)	Base	14/11	3/2	14/8	3/4
	Full SRAH	**15/14**	**1/1**	**15/13**	**1/2**
OpenVLA-ft Kim et al. (2025)	Base	16/13	3/2	15/10	3/3
	Full SRAH	**16/16**	**1/0**	**15/14**	**1/1**

5.3 Changes Localize to Diagnosed Failure Modes

Figure 3 localizes the changes: thin-shell simulation shows fewer environment-critical exposure events and robot-critical transport events; branching shows its largest reductions during transport. Physical branching trials likewise show fewer extraction events. These observations connect the aggregate gain to GRAM's diagnosis, without uniquely attributing each change to one module.

Table 2: **Component contributions across three backbones.** Simulation percentages are equal-weight means over Diffusion Policy, ACT-style, and OpenVLA-ft; E/R denotes Tightened-budget environment/robot violations. Per-backbone confidence intervals appear in Table 11. Physical counts pool 72 executions per analogue (24 per backbone). Both Δ columns report mean gentle-success gains in percentage points. The first non-base row adds both the residual head and stage conditioning.

	Thin-shell folds			Branching corridors			
Variant	Cross ↑	Tight ↑	E/R viol. ↓	Cross ↑	Tight ↑	E/R viol. ↓	Δ mean (pp) ↑
Base policy	37.1	41.8	25.3/22.7	44.1	44.9	20.7/19.0	+0.0
+ stage	39.6	43.9	23.8/21.5	45.2	46.2	19.6/18.0	+1.7
+ stage + budget	40.3	45.5	21.2/19.2	47.0	48.0	17.4/15.8	+3.2
+ stage + budget + hazard	40.2	46.5	19.8/17.8	47.4	49.1	16.2/14.5	+3.8
Full, gentle-only predictors	41.4	46.1	20.5/18.4	48.0	48.4	17.1/15.5	+4.0
Full SRAH	**42.3**	**47.4**	**19.1/16.9**	**49.1**	**50.1**	**15.3/13.5**	**+5.2**

	Thin-shell physical analogue		Branching physical analogue		
Variant	Raw/Gentle ↑	E/R viol. ↓	Raw/Gentle ↑	E/R viol. ↓	Δ mean (pp) ↑
Base policy	43/33	10/7	42/25	10/12	+0.0
+ stage	43/36	8/7	42/28	9/10	+4.2
+ stage + budget	43/39	5/4	42/33	6/7	+9.7
+ stage + budget + hazard	43/39	5/3	43/36	5/6	+11.8
Full SRAH	**45/42**	**4/2**	**44/38**	**4/6**	**+15.3**

Table 3: **Mechanics response and hazard provide complementary context.** Means over three backbones and both evaluation families. All rows include stage and budget; m_t indicates the dedicated mechanics encoder, not access to history elsewhere. E/R reports Tightened-budget violations.

Conditioning	ID?	m_t?	Cross ↑	Tight ↑	E/R viol. ↓
Stage + budget	No	No	43.7	46.8	19.3/17.5
+ family ID	Yes	No	44.0	47.1	18.9/17.2
+ mechanics latent	No	Yes	44.6	47.5	18.7/17.0
+ hazard	No	No	43.8	47.8	18.0/16.2
Mechanics + hazard	No	Yes	**45.7**	**48.7**	**17.2/15.2**

5.4 Learned Signals Track Interaction State

Stage accuracy is 86–95% across the evaluated regimes (Fig. 4). Environment/robot hazard AUROCs are 0.89/0.91, with average precision 0.60/0.64; mechanics predictions achieve $r = 0.67$–0.72 for continuous outcomes and AUROC 0.86/0.88 for attachment/snag. Before recorded crossings, hazard rises and the gate admits a smaller fraction of each proposal. These measurements support the functional, physical, and temporal interpretation of the learned state.

5.5 Physical Analogues Improve Gentle Retrieval

A Franka arm retrieves through an EVA-foam fold pocket and a coated-wire snag corridor, with 24 fixed configurations per backbone and analogue. Each backbone is physically fine-tuned before freezing; SRAH uses a disjoint training split (Appendix F). Across the three backbones, gentle completions increase from 33 to 42 of 72 thin-shell trials and from 25 to 38 of 72 branching trials, while raw completions increase only from 43 to 45 and 42 to 44. The larger gentle-success gains support improved retrieval acceptability in these controlled analogues, not zero-shot sim-to-real transfer.

6 Limitations

GRAM isolates two yielding-occluder families rather than spanning contact-rich manipulation. Its EAT factorization is specific to retrieval; insertion, tool use, and bimanual tasks require different

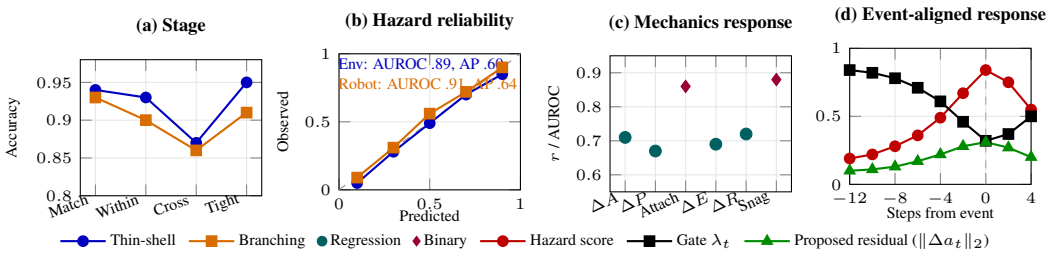

Figure 4: **Learned signals track functional, physical, and temporal state. (a)** Stage accuracy by regime. **(b)** Hazard discrimination and empirical reliability for environment (blue) and robot (orange). **(c)** Prediction of access (ΔA), progress (ΔP), attachment stability, environment/robot risk changes ($\Delta E/\Delta R$), and snag occurrence. **(d)** Around the first recorded budget crossing, hazard and residual proposals rise while the gate admits a smaller fraction of the proposal.

functional decompositions. The budgets quantify calibrated interaction-risk proxies, not material failure or certified safety. SRAH relies on current cost estimates, and its hazard predictions describe logged continuation behavior; as a learning-based approach, residual gating does not guarantee constraint satisfaction.

Our physical study uses two controlled analogues on one robot, with physical fine-tuning rather than zero-shot sim-to-real transfer. Resetting the media and replacing irreversibly damaged elements limits the scale of repeatable evaluation. The results establish improvements in these settings, not broad deployment readiness. Extending the diagnostic-to-adaptation approach requires testing more diverse materials, sensing conditions, and embodiments, including how errors in interaction-state estimates affect behavior.

7 Conclusion

GRAM makes gentle retrieval through yielding occlusions measurable through controlled mechanics shifts, functional-stage diagnostics, and separate environment and robot budgets. Its diagnostics show that failures are structured by retrieval function, local interaction response, and short-horizon event risk rather than by appearance or aggregate task progress alone. SRAH operationalizes these findings as a lightweight residual adapter for existing manipulation policies. It conditions action correction on predicted retrieval stage, mechanics response from recent interaction history, remaining budget headroom, and imminent hazard, while using unsafe and failed rollouts as predictive counterexamples. Across controlled simulation regimes and targeted physical analogues, this design improves gentle retrieval and reduces the stage- and risk-localized failures identified by GRAM. Together, these results support a diagnostic-to-adaptation approach: identify interaction factors associated with failure, then test whether exposing them to a policy improves behavior. This connects benchmark design to policy design, rather than treating evaluation and adaptation as separate problems.

AI use statement

In this work, we use the generative AI tools for latex formatting bugs and identifying grammar problems only. We have not used generative AI tools for the other tasks, and the rest of the required disclosure tasks are not applicable to this work. We have reviewed all AI-assisted work manually. We take responsibiliyt for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

Ethics statement

This work studies robotic interaction intended to reduce unnecessary disturbance to surrounding structures and excessive robot-side contact risk. As such, failures in real-robot experiments means breaking or damaging the physical analogues. This makes it important to use reusable, recyclable, and safe materials, alongside safe disposal protocols, so that this (often irreversible) process does not produce waste.

Reproducibility statement

We will release the code and data used in producing the manuscript upon publication.

References

Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. Constrained policy optimization. In *International conference on machine learning*, pp. 22–31. Pmlr, 2017.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.

Jiayuan Gu, Fanbo Xiang, Xuanlin Li, Zhan Ling, Xiqiang Liu, Tongzhou Mu, Yihe Tang, Stone Tao, Xinyue Wei, Yunchao Yao, Xiaodi Yuan, Pengwei Xie, Zhiao Huang, Rui Chen, and Hao Su. Maniskill2: A unified benchmark for generalizable manipulation skills. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=b_CQDy9vrD1.

Sanjay Haresh, Daniel Dijkman, Apratim Bhattacharyya, and Roland Memisevic. Clevrskills: Compositional language and visual reasoning in robotics. *Advances in Neural Information Processing Systems*, 37:38235–38266, 2024.

Zhiao Huang, Yuanming Hu, Tao Du, Siyuan Zhou, Hao Su, Joshua B. Tenenbaum, and Chuang Gan. Plasticinelab: A soft-body manipulation benchmark with differentiable physics. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=xCcdBRQEDW.

Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J Davison. Rlbench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, 5(2):3019–3026, 2020.

Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti, Soroush Nasiriany, Mohan Kumar Srirama, Lawrence Yunliang Chen, Kirsty Ellis, et al. Droid: A large-scale in-the-wild robot manipulation dataset. *arXiv preprint arXiv:2403.12945*, 2024.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan P Foster, Pannag R Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. Openvla: An open-source vision-language-action model. In Pulkit Agrawal, Oliver Kroemer, and Wolfram Burgard (eds.), *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 2679–2713. PMLR, 06–09 Nov 2025.

Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. Rma: Rapid motor adaptation for legged robots. *arXiv preprint arXiv:2107.04034*, 2021.

Vikash Kumar, Rutav Shah, Gaoyue Zhou, Vincent Moens, Vittorio Caggiano, Abhishek Gupta, and Aravind Rajeswaran. Robohive: A unified framework for robot learning. *Advances in Neural Information Processing Systems*, 36:44323–44340, 2023.

Jongmin Lee, Cosmin Paduraru, Daniel J Mankowitz, Nicolas Heess, Doina Precup, Kee-Eung Kim, and Arthur Guez. Coptidice: Offline constrained reinforcement learning via stationary distribution correction estimation. *arXiv preprint arXiv:2204.08957*, 2022.

Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Wensi Ai, Benjamin Martinez, et al. Behavior-1k: A human-centered, embodied ai benchmark with 1,000 everyday activities and realistic simulation. *arXiv preprint arXiv:2403.09227*, 2024.

Xingyu Lin, Yufei Wang, Jake Olkin, and David Held. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation. In *Conference on Robot Learning*, 2020.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning. *Advances in Neural Information Processing Systems*, 36:44776–44791, 2023a.

Zuxin Liu, Zijian Guo, Yihang Yao, Zhepeng Cen, Wenhao Yu, Tingnan Zhang, and Ding Zhao. Constrained decision transformer for offline safe reinforcement learning. In *International conference on machine learning*, pp. 21611–21630. PMLR, 2023b.

Haoran Lu, Ruihai Wu, Yitong Li, Sijie Li, Ziyu Zhu, Chuanruo Ning, Yan Shen, Longzan Luo, Yuanpei Chen, and Hao Dong. Garmentlab: A unified simulation and benchmark for garment manipulation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=bIRcf8i1kp.

Tongzhou Mu, Zhan Ling, Fanbo Xiang, Derek Yang, Xuanlin Li, Stone Tao, Zhiao Huang, Zhiwei Jia, and Hao Su. Maniskill: Generalizable manipulation skill benchmark with large-scale demonstrations. *arXiv preprint arXiv:2107.14483*, 2021.

Abby O'Neill, Abdul Rehman, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, et al. Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6892–6903. IEEE, 2024.

Wilbert Pumacay, Ishika Singh, Jiafei Duan, Ranjay Krishna, Jesse Thomason, and Dieter Fox. The colosseum: A benchmark for evaluating generalization for robotic manipulation. *arXiv preprint arXiv:2402.08191*, 2024.

Kate Rakelly, Aurick Zhou, Chelsea Finn, Sergey Levine, and Deirdre Quillen. Efficient off-policy meta-reinforcement learning via probabilistic context variables. In *International conference on machine learning*, pp. 5331–5340. PMLR, 2019.

Daniel Seita, Pete Florence, Jonathan Tompson, Erwin Coumans, Vikas Sindhwani, Ken Goldberg, and Andy Zeng. Learning to rearrange deformable cables, fabrics, and bags with goal-conditioned transporter networks. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4568–4575. IEEE, 2021.

Bokui Shen, Zhenyu Jiang, Christopher Choy, Silvio Savarese, Leonidas J Guibas, Anima Anandkumar, and Yuke Zhu. Action-conditional implicit visual dynamics for deformable object manipulation. *The International Journal of Robotics Research*, 43(4):437–455, 2024.

Mohit Shridhar, Lucas Manuelli, and Dieter Fox. Perceiver-actor: A multi-task transformer for robotic manipulation. In *Conference on Robot Learning*, pp. 785–799. PMLR, 2023.

Neha Sunil, Shaoxiong Wang, Yu She, Edward Adelson, and Alberto Rodriguez Garcia. Visuotactile affordances for cloth manipulation with local control. In *6th Annual Conference on Robot Learning*, 2022. URL https://openreview.net/forum?id=s6NEzqZKaP-.

Andrew Szot, Alexander Clegg, Eric Undersander, Erik Wijmans, Yili Zhao, John Turner, Noah Maestre, Mustafa Mukadam, Devendra Singh Chaplot, Oleksandr Maksymets, et al. Habitat 2.0: Training home assistants to rearrange their habitat. *Advances in neural information processing systems*, 34:251–266, 2021.

Homer Rich Walke, Kevin Black, Tony Z Zhao, Quan Vuong, Chongyi Zheng, Philippe Hansen-Estruch, Andre Wang He, Vivek Myers, Moo Jin Kim, Max Du, et al. Bridgedata v2: A dataset for robot learning at scale. In *Conference on Robot Learning*, pp. 1723–1736. PMLR, 2023.

Ruihai Wu, Chuanruo Ning, and Hao Dong. Learning foresightful dense visual affordance for deformable object manipulation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10947–10956, 2023.

Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning. In *Conference on robot learning*, pp. 1094–1100. PMLR, 2020.

Andy Zeng, Pete Florence, Jonathan Tompson, Stefan Welker, Jonathan Chien, Maria Attarian, Travis Armstrong, Ivan Krasin, Dan Duong, Vikas Sindhwani, et al. Transporter networks: Rearranging the visual world for robotic manipulation. In *Conference on Robot Learning*, pp. 726–747. PMLR, 2021.

Tony Z Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

A Appendix

In this appendix section, we are provide further details on the benchmark setup, the validity of the metrics proposed and used, the implementation of the proposed SRAH method and further supporting results, the physical experiment protocols, fairness and robustness insighs to the benchmark, and lastly further secondary result of interest.

B Benchmark Specification

This section specifies the task semantics, procedural parameters, evaluation splits, and corpus sizes used by GRAM. The benchmark is designed to vary the mechanics of yielding-occlusion retrieval while keeping the robot interface, retrieval objective, and budget semantics fixed.

B.1 Task Semantics and EAT Attribution

All tasks use the same retrieval interface and the same budget semantics as the main paper. A rollout is a gentle success only if the target reaches the goal and both normalized environment-disturbance and robot self-preservation costs remain within budget. The Expose–Attach–Transport decomposition is used for diagnostic attribution and stage supervision; it does not impose a scripted execution order.

We define an access predicate A_t and an attachment predicate H_t. In simulation,

$$A_t = 1 \tag{3}$$

when target visibility exceeds a fixed access threshold and the robot has a collision-free approach corridor to the target. The visibility threshold is held fixed across experiments. The attachment predicate satisfies

$$H_t = 1 \tag{4}$$

when the gripper is in contact with the target and target motion remains consistent with gripper motion over a short temporal window. Stage labels give stable attachment precedence over access:

$$z_t = \begin{cases} \textsc{Transport}, & H_t = 1, \\ \textsc{Attach}, & H_t = 0 \wedge A_t = 1, \\ \textsc{Expose}, & H_t = 0 \wedge A_t = 0. \end{cases} \tag{5}$$

Thus an acquired target remains in Transport when it becomes re-occluded. Predicate changes must persist for three consecutive control steps before the stage label changes. Predicate changes must persist for three consecutive control steps before the stage label changes, reducing label flicker near access or attachment boundaries.

Visibility is computed as the fraction of target surface area directly visible under depth-based occlusion reasoning across the calibrated camera set. Accessibility is computed from a library of $K = 48$ candidate pregrasp poses distributed around the target at a 4–6 cm standoff. Candidates are filtered by inverse-kinematics feasibility and collision-free approach, and accessibility is the best available approach-corridor clearance:

$$\mathcal{A} = \max_{k \in \mathcal{F}} \text{clearance}(k), \tag{6}$$

where $\mathcal{F}$ is the set of feasible candidate pregrasps.

Unless otherwise noted, tasks use a horizon of $T = 300$ steps and a goal-dwell requirement of 10 steps. The goal region is a sphere of radius 4 cm located outside the occluder volume. For exposure-only diagnostics, the interaction horizon is 200 steps.

Table 4: **GRAM task suite.** All tasks additionally require compliance with the environment and robot budgets. The Expose, Attach, and Transport labels denote diagnostic function rather than a mandatory state-machine order.

Task	Stage focus	Family	Initialization	Objective / termination
Occluder Management (Contact-Free)	Expose	Both	Target occluded; grasp initially infeasible	Improve visibility and accessibility while limiting disturbance. Target contact is a failure for this diagnostic only; terminate on contact, timeout, or catastrophic robot event.
Occluder Management (Contact-Allowed)	Expose	Both	Same as above	Same exposure objective, but incidental target contact is permitted and logged.
Attachment Under Full Visibility	Attach	Both	Target visible and reachable; occluder constrains approach	Establish stable attachment and deliver to the goal under both budgets.
Attachment Under Partial Visibility	Attach	Both	Target intermittently visible; approach may be occluded	Establish stable attachment and deliver to the goal; interaction may first improve visibility or access.
Carry in Free Space	Transport	Both	Begins with verified stable attachment and nominally clear path	Preserve attachment and reach the goal while respecting robot self-preservation constraints.
Carry Through a Yielding Corridor	Transport	Both	Begins attached; contact with occluder is unavoidable	Reach the goal while limiting thin-shell disturbance or branch snag/impulse escalation.
Reach-Through Retrieval	Composite	Both	Target initially occluded	End-to-end retrieval with arbitrary exposure/acquisition interleavings and regrasping permitted.
Fold-Pocket Retrieval	Composite	Thin-shell	Target fully hidden in a deep fold	End-to-end retrieval emphasizing contact-mediated exposure while avoiding concentrated fold-mouth disturbance.
Snag-Aware Retrieval	Composite	Branching	Target often partly visible; extraction is snag-prone	End-to-end retrieval emphasizing retreat, replanning, and low-impulse extraction.
Time-Varying Occlusion Retrieval	Composite	Thin-shell	Occluder evolves through controlled growth/relaxation	End-to-end retrieval under changing access corridors and contact opportunities.

For thin-shell tasks, the target is a rigid sphere of diameter 3 cm placed near a fold pocket by sampling on the shell and offsetting inward along the local normal. Placements with initial interpenetration are rejected, with at least 2 mm initial shell clearance. For branching tasks, target positions are sampled inside the branch volume and rejected if the initial corridor clearance is too small. Target mass is randomized by $\pm 10\%$.

Access predicate. The access predicate is combine the target observability with executable reachability. Let $V_t \in [0, 1]$ denote the fraction of target surface visible from the calibrated camera set, and let $\mathcal{F}_t$ denote the subset of the $K = 48$ candidate pregrasp poses that are inverse-kinematically feasible and admit a collision-free swept approach to the target. So, we define

$$A_t = \mathbb{1}\left[V_t \geq \tau_V \ \wedge \ |\mathcal{F}_t| > 0\right], \qquad \tau_V = 0.30. \tag{7}$$

The visibility threshold is fixed globally and is not varied across occluder families, tasks, methods, or evaluation regimes. The criterion is intentionally permissive enough to support attachment under partial visibility while rejecting brief or incidental target glimpses that do not provide actionable access. To suppress stage flicker, a predicate change must persist for three consecutive control steps before changing the stage label.

We verified during benchmark calibration that the chosen threshold lies well above visibility levels produced by incidental sliver observations while remaining below the visibility typically available at successful first attachment.

B.2 Procedural Families and Generator Parameters

The procedural families are used as controlled interventions on interaction mechanics rather than as a visual-diversity benchmark. Thin-shell folds vary pocket geometry and material response; branching corridors vary corridor density, tortuosity, and clearance. Table 5 gives the exact parameterization.

Derived quantities such as visibility, accessibility, and entanglement risk are used for auditing and stratified analysis, not as constraints on policy behavior. In particular, the branching generator exposes

Table 5: **Primary generator knobs and difficulty tiers.** Knobs are sampled independently within the listed tier intervals.

Family	Knob	Semantics	Base range	Distribution	Easy	Medium	Hard
Thin-shell	Fold severity ϕ	Pocket depth, curvature, and fold-mouth narrowing	$[0, 1]$	$\mathrm{Beta}(2, 2)$	$[0.18, 0.39]$	$[0.39, 0.61]$	$[0.61, 0.82]$
Thin-shell	Compliance scale k	In-plane and bending compliance	$[0.6, 1.6]$	LogUniform	$[0.69, 0.86]$	$[0.86, 1.12]$	$[1.12, 1.39]$
Thin-shell	Heterogeneity h	Spatial stiffness variation	$[0, 0.5]$	Uniform	$[0.00, 0.12]$	$[0.12, 0.32]$	$[0.32, 0.50]$
Thin-shell	Dynamics rate ω	Growth/relaxation rate for dynamic occlusion	$[0, 1]$	$\mathrm{Beta}(1.5, 3)$	$[0.00, 0.20]$	$[0.20, 0.45]$	$[0.45, 0.75]$
Branching	Density δ	Branch count per volume and corridor formation	$[0.5, 1.5]$	scaled $\mathrm{Beta}(2, 2)$	$[0.70, 0.92]$	$[0.92, 1.08]$	$[1.08, 1.30]$
Branching	Tortuosity ν	Curvature and hook-like contact frequency	$[0, 1]$	$\mathrm{Beta}(2.5, 2)$	$[0.28, 0.48]$	$[0.48, 0.66]$	$[0.66, 0.84]$
Branching	Radius scale r	Effective clearance and compliance	$[0.7, 1.3]$	Uniform	$[0.70, 0.88]$	$[0.88, 1.12]$	$[1.12, 1.30]$

a radius-annotated skeleton graph, from which a maximum-clearance nominal extraction path is computed. The entanglement-risk diagnostic combines narrow-passage clearance and predicted near-contact opportunities along this path.

B.3 Evaluation Regimes

All learned policies are trained once on the matched training distribution and evaluated without adaptation unless explicitly stated otherwise.

Matched. Train and test morphologies are drawn from the same central generator support.

Within-family morphology shift. The occluder family is fixed while mechanics-relevant generator parameters move away from the training support. Let F_j denote the base CDF of generator knob j. Training uses the central quantile interval $[0.35, 0.65]$. At shift magnitude $s \in [0, 1]$, the evaluation interval is

$$[q_s^-, q_s^+] = (1 - s)[0.35, 0.65] + s[0.85, 0.95], \tag{8}$$

after which

$$u \sim \mathrm{Uniform}(q_s^-, q_s^+), \qquad \theta_j = F_j^{-1}(u). \tag{9}$$

For thin shells, the principal shift coordinates are fold severity and stiffness heterogeneity; for branching corridors they are density, tortuosity, and radius scaling.

Cross-family transfer. We fit the policy on one occluder family and evaluate on the other; reported columns identify the evaluation family. All benchmark-specific fitting of the backbone and adapter—including feature and target normalization, calibration of predictive scores, and model selection—uses source-family training or validation data only. The target family supplies no policy or adapter fitting data. Benchmark metric scales are calibrated once per family, independently of policy fitting, as specified in Appendix C.3.

Tightened budget. Morphology and task semantics are unchanged, while both nominal budgets are multiplied by $s_b = 0.80$. Raw cost definitions, component normalization, and composite weights remain fixed. Policies are executed under the reduced budgets without parameter updates. Budget-conditioned inputs, including SRAH's headroom, use the reduced limits; gentle-success labels and hazard-evaluation targets use the same thresholds.

Capture shift. Camera pose, lighting, background appearance, and mild calibration noise are perturbed while morphology and task mechanics are held fixed. This provides a control for appearance-driven distribution shift.

Observation ladder. Diagnostic variants modify available sensing while leaving actions and evaluation unchanged. These variants are used to distinguish information limitations from strategy limitations and are not part of the main headline comparison.

Table 6: **GRAM corpus size.**

Family	Split	Episodes	Median steps	Frames (M)	Sim. hours
Thin-shell	Train	68,000	190	12.9	179
Thin-shell	Val.	8,000	190	1.5	21
Thin-shell	Test	10,000	200	2.0	28
Branching	Train	72,000	175	12.6	175
Branching	Val.	8,000	175	1.4	19
Branching	Test	10,000	185	1.9	26
Total		176,000		32.3	448

B.4 Corpus and Evaluation Counts

The complete corpus contains 176,000 trajectories, 32.3 million logged frames, and 448 simulated hours. Evaluation uses fixed episode sets shared across methods, enabling paired comparisons and stratified bootstrap confidence intervals.

Each family contributes 2,000 fixed evaluation episodes to each of the five main regimes: Matched, Within-family morphology shift, Cross-family transfer, Tightened budget, and Capture shift. Confidence intervals are computed by stratified bootstrap over episodes with 10,000 resamples.

C Metrics and Construct Validity

This section answers two questions about the benchmark costs: how they are computed, and whether the resulting quantities track interaction outcomes that are meaningfully undesirable. We treat EnvDist and RobotCost as calibrated interaction-risk proxies rather than material-failure models.

C.1 Environment Disturbance

Environment disturbance combines sustained deformation, localized peak events, and overall structural displacement:

$$D_{\text{env}} = \alpha D_{\text{def}} + \beta D_{\text{peak}} + \gamma D_{\text{disp}}, \qquad \alpha + \beta + \gamma = 1. \tag{10}$$

For thin-shell environments,

$$D_{\text{def}} = \sum_{t=1}^{T} \left(E_{\text{strain}}(t) + \lambda_b E_{\text{bend}}(t) \right), \tag{11}$$

where E_{strain} and E_{bend} are simulator-logged membrane and bending energies. The peak term is the maximum vertex-local energy density,

$$D_{\text{peak}} = \max_{t \le T} \max_{v \in V} \frac{E_v(t)}{A_v}, \tag{12}$$

where A_v is the mixed Voronoi/barycentric area associated with vertex v. Structural displacement is

$$D_{\text{disp}} = \sum_{t=1}^{T} \sum_{v \in V} A_v \left\| \Delta x_v(t) \right\|_2^2, \tag{13}$$

after removing rigid motion of the anchored support.

For branching environments, sustained deformation uses the generator's radius-annotated skeleton:

$$D_{\text{def}} = \sum_{t=1}^{T} \sum_{e \in E} \left(k_b r_e^4 \kappa_e(t)^2 + k_\tau r_e^4 \tau_e(t)^2 \right), \tag{14}$$

with $k_b = k_\tau = 1$, local radius r_e, discrete curvature κ_e, and torsion proxy τ_e. The peak term is the maximum localized segment contribution over time. Displacement is

$$D_{\text{disp}} = \sum_{t=1}^{T} \sum_{p \in P} r_p^2 \left\| \Delta x_p(t) \right\|_2^2, \tag{15}$$

Table 7: **Calibration anchors.** Entries are median / p95 / violation rate (%).

Family	Controller	EnvDist	RobotCost
Thin-shell	Conservative	0.98 / 1.18 / 8	0.95 / 1.14 / 6
Thin-shell	Destructive	1.62 / 2.35 / 88	1.55 / 2.20 / 84
Branching	Conservative	0.97 / 1.20 / 9	0.96 / 1.16 / 7
Branching	Destructive	1.58 / 2.28 / 86	1.62 / 2.34 / 87

where the r_p^2 weighting approximates mass-per-length weighting and prevents thin peripheral branches from dominating displacement.

C.2 Robot Self-Preservation

Robot-side cost combines sustained load, abrupt load variation, and contact impulse:

$$C_{\text{robot}} = \eta \sum_{t=1}^{T} \|\hat{\tau}(t)\|_2^2 + \kappa \sum_{t=2}^{T} \|\hat{\tau}(t) - \hat{\tau}(t-1)\|_2^2 + \rho \sum_{c \in \mathcal{C}} \hat{J}_c. \tag{16}$$

Joint torques are normalized by per-joint torque limits. Each accumulated contact impulse is normalized by

$$\hat{J}_c = \frac{J_c}{m_{\text{tgt}} v_{\text{ref}}}, \qquad v_{\text{ref}} = \omega_{\text{ctrl}} \ell_{\text{ref}}, \tag{17}$$

where ℓ_{ref} is the nominal target diameter. Self-collision and joint-limit events are treated as hard safety failures.

The normalized episode costs are

$$\bar{D}_{\text{env}} = \frac{D_{\text{env}}}{B_{\text{env}}}, \qquad \bar{C}_{\text{robot}} = \frac{C_{\text{robot}}}{B_{\text{robot}}}, \tag{18}$$

with budget violation at values greater than one.

C.3 Calibration

Calibration uses 2,000 held-out composite episodes per family from the matched morphology distribution. Two scripted controllers establish the operating range: a conservative controller using low-impedance, bounded-speed, retreat-on-contact behavior, and an intentionally destructive stress test using high speed, high stiffness, no retreat, and direct pushing.

For each constituent raw term z, we compute

$$\tilde{z} = \frac{z}{q_{0.9}(z)}, \tag{19}$$

where $q_{0.9}$ is the empirical 90th percentile under the conservative anchor. Composite weights are then selected by variance equalization over the scaled terms. The values used in all experiments are

$$(\alpha, \beta, \gamma) = (0.50, 0.30, 0.20), \qquad \lambda_b = 0.35, \tag{20}$$

and

$$(\eta, \kappa, \rho) = (0.55, 0.30, 0.15). \tag{21}$$

The normalized nominal budgets are $B_{\text{env}} = B_{\text{robot}} = 1$.

C.4 Construct Validity Against Human-Labeled Events

We evaluate whether the disturbance proxies correspond to visually recognizable undesirable interaction. Human labels are collected on 2–4 s clips sampled using a mixture of proxy-peak windows, uniformly random windows, and low-EnvDist windows. This avoids constructing the evaluation set solely from events selected by the same proxy being validated.

Table 8: **Construct validity of the disturbance proxies.** Panel (a) reports ROC–AUC for predicting blinded human-labeled damage-like clips ($N = 2{,}000$ per family). Panel (b) reports the episode-level rate of human-labeled events below and above the EnvDist budget. Panel (c) tests whether method ordering depends strongly on the particular disturbance definition.

(a) Human-label prediction

Family	Proxy	AUC
Thin-shell	EnvDist composite	0.87±0.02
Thin-shell	Peak energy density	0.90±0.02
Thin-shell	Displacement only	0.74±0.03
Thin-shell	Max impulse	0.78±0.03
Branching	EnvDist composite	0.84±0.02
Branching	Peak bend/torsion	0.86±0.02
Branching	Displacement only	0.70±0.03
Branching	Max impulse	0.82±0.02

(b) Budget boundary

Region	Thin	Branch
EnvDist ≤ 1	0.11	0.14
EnvDist > 1	0.48	0.52

(c) Alternative metrics

Family	Metric	τ
Thin	Displacement	0.84
Thin	Max impulse	0.79
Thin	Duration $>$ threshold	0.86
Branch	Displacement	0.80
Branch	Max impulse	0.88
Branch	Duration $>$ threshold	0.83

Table 9: **Ranking stability under metric perturbations.** Entries are mean / minimum / maximum Kendall τ relative to the default configuration, reported for success ranking.

Sweep	Thin-shell	Branching
EnvDist weights	0.86 / 0.74 / 0.93	0.84 / 0.71 / 0.92
RobotCost weights	0.89 / 0.81 / 0.95	0.88 / 0.80 / 0.95
Budget scale	0.91 / 0.84 / 0.96	0.90 / 0.83 / 0.95

Three blinded annotators independently mark whether each clip contains a damage-like event. Thin-shell positives include pinch/tear-like stress or persistent localized fold-mouth concentration; branching positives include snag escalation with sudden release or whip-like recoil; hard collisions count as robot-side risk in both families. Fleiss' κ on pre-adjudication binary labels is 0.72, and 7% of clips are marked uncertain. Final labels use majority vote with adjudication only when necessary.

These results support the intended interpretation of EnvDist as an interaction-risk proxy: the composite and peak-sensitive quantities discriminate human-labeled events substantially better than displacement alone, episodes beyond the budget contain labeled events much more frequently, and benchmark conclusions are not tied to a single disturbance definition.

C.5 Sensitiviyt to Weights and Budgets

We additionally test whether method rankings are sensitive to reasonable changes in the metric definition. Disturbance weights (α, β, γ) and robot weights (η, κ, ρ) are swept over their simplices, while the common budget scale is varied over

$$s_b \in \{0.70, 0.80, 0.90, 1.00, 1.10, 1.20, 1.30\}. \tag{22}$$

The benchmark conclusions are therefore stable across substantial perturbations of both composite weights and budget thresholds rather than depending on one narrow calibration choice.

D SRAH Implementation

This section specifies the architecture, auxiliary targets, loss functions, data partitions, and inference procedure used by SRAH.

D.1 Inputs and Architecture

All SRAH variants operate on the same normalized Cartesian action space as their frozen base policy: end-effector translation delta, rotation delta, and gripper command. The adapter receives the base action proposal a_t^0, recent interaction history over $L = 8$ control steps, normalized budget headroom, and backbone features when available. If the base policy does not expose internal features, the adapter uses its own observation-history embedding.

Table 10: **SRAH architecture and training hyperparameters.**

Component	Configuration
Base policy	Frozen throughout SRAH training
Control rate	10 Hz
Action space	Normalized Cartesian translation delta, rotation delta, gripper command
History length L	8 control steps
Prediction horizon H	12 control steps
Stage estimator q_ϕ	2-layer MLP, hidden width 256, SiLU
Mechanics encoder e_η	1-layer GRU, hidden size 64, projected to 32-D latent
Mechanics prediction head p_ξ	2-layer MLP, hidden width 256
Hazard critic r_ψ	2-layer MLP, hidden width 256, two sigmoid outputs
Residual head g_θ	2-layer MLP, hidden width 256, FiLM conditioning by stage posterior
Gate	Learned affine-sigmoid; $\alpha_0 = 0$, $\alpha_b = 2$, $\alpha_\rho = 2$ initialization
Optimizer	AdamW
Learning rate	3×10^{-4}
Weight decay	10^{-4}
Batch size	256 timesteps
Epochs	50
Gradient clipping	Global norm 1.0
Hazard class balancing	Per-minibatch positive-class reweighting, clipped at 10
Loss weights	$\lambda_z = 1.0$, $\lambda_m = 0.5$, $\lambda_h = 1.0$, $\lambda_s = 10^{-3}$
Residual supervision	Gentle successful rollouts only
Predictive supervision	Gentle, unsafe, and failed rollouts

The recent history contains executed actions, proprioception, gripper state, target-relative motion features when available, and contact or force/torque proxies.

All prediction inputs are causal. At control step t, h_t contains observations available through t and previously executed actions through $t-1$; future changes appear only as training targets. Budget headroom uses current cost telemetry: calibrated simulator costs in simulation, and tracked structural deformation plus robot telemetry on hardware. Benchmark access and attachment predicates supervise the stage estimator; the residual conditions on its predicted posterior.

D.2 Stage and Mechanics-Response Targets

Stage supervision uses Eq. 5. These labels are used only to train the stage estimator; at deployment the model conditions on the predicted posterior $\hat{z}_t$ and does not execute a hand-coded state machine.

The mechanics-response latent is trained to predict short-horizon interaction consequences over $H = 12$ steps:

$$u_t = \left[\Delta\text{access}, \Delta\text{progress}, \text{stable}, \Delta D^E, \Delta D^R, \text{snag}\right]_{t:t+H}. \tag{23}$$

Here Δaccess is the change in the visibility/reachability access score, $\Delta\text{progress}$ is the change in normalized target-to-goal progress, stable indicates whether attachment remains stable, ΔD^E and ΔD^R are increases in normalized environment and robot risk, and snag indicates that target progress stalls while contact load or occluder displacement increases.

For predictive supervision, the mechanics head conditions on the logged action associated with the observed future:

$$\hat{u}_t^{\log} = p_\xi(m_t, \hat{z}_t, a_t^{\log}). \tag{24}$$

Here $a_t^{\log}$ is the action executed at the start of the target window. The remaining actions in that window follow the recorded continuation behavior. Thus the auxiliary task predicts observed interaction consequences, rather than assigning those consequences to an unexecuted base-policy proposal. The

mechanics-response loss is

$$\begin{aligned}\mathcal{L}_{\text{mech}} = &\left|\widehat{\Delta\text{access}} - \Delta\text{access}\right|_1 + \left|\widehat{\Delta\text{progress}} - \Delta\text{progress}\right|_1 \\ &+ \left|\widehat{\Delta D^E} - \Delta D^E\right|_1 + \left|\widehat{\Delta D^R} - \Delta D^R\right|_1 \\ &+ \text{BCE}\left(\widehat{\text{stable}}, \text{stable}\right) + \text{BCE}\left(\widehat{\text{snag}}, \text{snag}\right).\end{aligned} \tag{25}$$

Continuous targets are normalized using training-set statistics.

D.3 Hazard Supervision

Let $D_t^E = \bar{D}_{\text{env},t}$ and $D_t^R = \bar{C}_{\text{robot},t}$ use the active training or evaluation budgets. For channel $k \in \{E, R\}$, define its first crossing time $T^k = \inf\{s : D_s^k > 1\}$. Valid pre-crossing targets are

$$y_t^k = \mathbb{1}[t < T^k \le t + H], \qquad t < T^k, \qquad H = 12. \tag{26}$$

Post-crossing samples are excluded for that channel. An observed crossing determines a positive label even when the remainder of the horizon is unavailable. A negative label requires a fully observed horizon or an absorbing task termination with no crossing; other incomplete windows are masked out. Channel-specific masks allow prediction of the other channel to continue after one channel crosses.

During predictive training,

$$\hat{\rho}_t^{\log} = r_\psi(h_t, a_t^{\log}, \hat{z}_t, m_t, b_t), \tag{27}$$

where $a_t^{\log}$ is the executed action associated with the future label. Binary cross-entropy uses per-minibatch positive-class reweighting capped at 10 and is averaged over valid channel–timestep pairs. The predictions summarize crossing risk under logged continuation behavior. At deployment, the head is queried with a_t^0; this does not identify the counterfactual crossing probability after the residual changes that action. Because class-weighted training does not itself guarantee calibration, we evaluate discrimination and empirical reliability separately.

D.4 Residual Adapter and Gate

Continuous commands are normalized to $[-1, 1]$ and executed using Eq. 1. The gate is

$$\lambda_t = \sigma\left(\alpha_0 + \alpha_b \min_{k\in\{E,R\}} b_t^k - \alpha_\rho \max_{k\in\{E,R\}} \hat{\rho}_t^k\right). \tag{28}$$

Initialization uses $\alpha_0 = 0$, $\alpha_b = 2$, and $\alpha_\rho = 2$. This favors smaller initial corrections near depleted budgets or predicted crossings; it does not constrain the backbone to stop or provide a safety certificate.

Residual imitation uses the unprojected *gated* proposal in Eq. 2; clipping is applied at execution. Both the residual head and gate receive gradients from this loss, while π_0 remains frozen. For this action-imitation branch, hazard is evaluated at a_t^0; predictive supervision instead uses $a_t^{\log}$. The regularizer is

$$\mathcal{L}_{\text{reg}} = \|\Delta a_t\|_2^2 + 0.5\|\Delta a_t - \Delta a_{t-1}\|_2^2. \tag{29}$$

Temporal pairs are adjacent control steps from the same trajectory. The penalties regularize residual proposals; action clipping bounds the executed command rather than imposing a hard bound on the residual itself.

D.5 Data Partitions and Training Objective

Logged trajectories are partitioned into

$$\mathcal{D}_{\text{all}} = \mathcal{D}_{\text{gentle}} \cup \mathcal{D}_{\text{unsafe}} \cup \mathcal{D}_{\text{fail}}. \tag{30}$$

$\mathcal{D}_{\text{gentle}}$ contains trajectories that reach the goal while remaining within both budgets. $\mathcal{D}_{\text{unsafe}}$ contains trajectories that reach the goal but violate at least one budget. $\mathcal{D}_{\text{fail}}$ contains trajectories that do not reach the goal.

Algorithm 1 SRAH inference at one control step

1: Form causal history h_t and update active-budget headroom b_t.
2: Obtain $(a_t^0, \mathcal{C}_{t+1}) = \mathrm{NextAction}(\pi_0, h_t, \mathcal{C}_t)$.
3: Predict stage posterior $\hat{z}_t = q_\phi(h_t)$.
4: Encode mechanics response $m_t = e_\eta(h_t)$.
5: Predict hazards $\hat{\rho}_t = r_\psi(h_t, a_t^0, \hat{z}_t, m_t, b_t)$.
6: Predict residual $\Delta a_t = g_\theta(h_t, a_t^0, \hat{z}_t, m_t, b_t, \hat{\rho}_t)$.
7: Compute λ_t using Eq. 28.
8: Execute $a_t = \Pi_{\mathcal{A}}(a_t^0 + \lambda_t \Delta a_t)$.
9: Store the executed a_t, rather than a_t^0, for subsequent history.

Only $\mathcal{D}_{\mathrm{gentle}}$ supervises residual imitation. Stage, mechanics-response, and hazard losses use $\mathcal{D}_{\mathrm{all}}$, allowing unsafe and failed interaction to supervise prediction without making those actions imitation targets.

The complete objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{res}} + \lambda_z \mathcal{L}_{\mathrm{stage}} + \lambda_m \mathcal{L}_{\mathrm{mech}} + \lambda_h \mathcal{L}_{\mathrm{haz}} + \lambda_s \mathcal{L}_{\mathrm{reg}}. \tag{31}$$

D.6 Inference

The native backbone wrapper retains its Base-only query cadence, action chunking, decoding, and temporal aggregation. Its cache $\mathcal{C}_t$ provides the next command, which SRAH modifies at each control step.

E Extended SRAH Results

This section expands the main-paper SRAH results to test whether the component trends hold separately across policy backbones and to specify the diagnostic evaluation used for the learned interaction-state heads.

E.1 Per-Backbone Component Ablations

Table 11 separates the three backbones averaged in the main paper. Full SRAH improves gentle success in every reported family–regime comparison. The first ablation increment introduces a residual adapter together with stage conditioning, so it does not isolate stage information from added adapter capacity. Subsequent comparisons vary the retained interaction-state signals; hazard gains need not be positive in every transfer cell.

Removing the mechanics encoder removes the explicit response latent m_t, not history already consumed by other retained components. The gentle-only predictor variant preserves the architecture and residual action targets but excludes unsafe and failed trajectories from predictive supervision. It therefore changes predictive data volume and hazard-event coverage, rather than providing a data-size-matched test of counterexamples. Its lower performance supports including risk-bearing interaction in predictive training.

Aggregation. Main-text simulation aggregates give equal weight to each of the three backbones. Table 2 preserves the four evaluation-family and regime cells; its mean gain averages their differences from Base. Table 3 additionally averages the two evaluation families. Aggregate tables report means, while per-backbone tables retain the reported confidence intervals. Physical aggregates sum counts over 24 executions per backbone, giving 72 executions per analogue and variant. Their mean percentage-point gain is $100[(g_{\mathrm{thin}} - g^0_{\mathrm{thin}}) + (g_{\mathrm{branch}} - g^0_{\mathrm{branch}})]/(2 \times 72)$. Rounding is applied only after aggregation.

E.2 Interaction-State Diagnostic Protocol

Figure 4 evaluates predictive state without selecting only successful trajectories. Prediction inputs are causal; future outcomes are used only as labels. Table 12 lists the reported quantities.

Table 11: **Per-backbone SRAH component ablations.** Gentle success (%) is reported with 95% stratified-bootstrap confidence intervals. Violation columns report Tightened-budget environment/robot rates. The first non-base row introduces both residual adaptation and stage conditioning.

Variant	Thin-shell folds			Branching corridors		
	Cross ↑	Tight ↑	E/R viol. ↓	Cross ↑	Tight ↑	E/R viol. ↓
Diffusion Policy						
Base policy	31.7±1.8	27.9±1.7	36.0/32.0	35.4±2.0	30.7±1.9	32.0/30.0
+ stage	34.4±1.9	30.4±1.8	33.8/30.5	36.8±1.9	32.2±1.9	30.4/28.4
+ stage + budget	35.1±1.9	32.6±1.8	30.5/27.5	38.8±1.9	34.4±1.8	27.8/25.9
+ stage + budget + hazard	34.9±1.9	34.0±1.8	28.8/25.7	39.2±1.9	36.0±1.9	25.9/23.6
Full, gentle-only predictors	36.5±1.9	33.5±1.8	29.7/26.4	40.3±1.9	35.1±1.9	26.8/24.8
Full SRAH	**37.8±1.9**	**35.0±1.8**	**27.8/24.6**	**41.5±1.9**	**37.4±1.9**	**24.7/22.1**
ACT-style						
Base policy	35.9±1.9	45.1±2.0	22.0/20.0	44.9±2.0	49.8±2.0	18.0/16.0
+ stage	38.6±1.9	47.6±1.9	20.0/18.5	46.2±1.9	51.1±1.9	16.2/14.6
+ stage + budget	39.5±1.8	48.8±1.9	18.0/16.5	47.7±1.9	53.1±1.9	14.5/12.5
+ stage + budget + hazard	39.0±1.8	49.6±1.9	17.0/15.6	48.3±1.9	54.0±1.9	13.0/11.2
Full, gentle-only predictors	40.2±1.8	49.8±1.9	17.2/15.6	48.8±1.9	53.3±1.9	13.8/12.2
Full SRAH	**41.2±1.8**	**50.8±1.9**	**16.2/14.5**	**49.4±1.9**	**54.4±1.9**	**12.2/10.7**
OpenVLA-ft						
Base policy	43.8±2.0	52.3±2.0	18.0/16.0	52.0±2.0	54.3±2.0	12.0/11.0
+ stage	45.8±1.9	53.6±1.9	17.5/15.5	52.6±1.9	55.3±1.9	12.2/11.0
+ stage + budget	46.4±1.9	55.1±1.9	15.0/13.5	54.6±1.9	56.5±1.9	10.0/9.0
+ stage + budget + hazard	46.6±1.9	56.0±1.9	13.6/12.2	54.7±1.9	57.4±1.9	9.8/8.6
Full, gentle-only predictors	47.4±1.9	55.0±1.9	14.6/13.2	55.0±1.9	56.7±1.9	10.6/9.6
Full SRAH	**48.0±1.9**	**56.4±1.9**	**13.2/11.5**	**56.4±1.9**	**58.4±1.9**	**9.0/7.7**

Stage and mechanics. Stage accuracy is reported separately for Matched, Within-family, Cross-family, and Tightened-budget evaluation. Continuous consequence predictions are evaluated with Pearson correlation; attachment stability and snag predictions use ROC–AUC.

Hazard. Each channel is evaluated on valid pre-crossing windows, using the same horizon and censoring rules as Appendix D.3. We report ROC–AUC, average precision, and empirical reliability curves comparing binned scores with observed crossing frequencies. These quantities distinguish ranking quality from probability calibration.

Event-aligned response. Histories are aligned to the first recorded environment or robot crossing. The figure reports mean hazard, gate value, and proposed residual norm from 12 steps before through four steps after the event, using observed samples. Post-event points describe adapter response, not first-crossing prediction accuracy. The plotted means do not determine the mean admitted correction:

$$\mathbb{E}[\|\lambda_t \Delta a_t\|_2] = \mathbb{E}[\lambda_t \|\Delta a_t\|_2] \neq \mathbb{E}[\lambda_t]\,\mathbb{E}[\|\Delta a_t\|_2] \tag{32}$$

in general. Event alignment characterizes behavior on crossing histories; it does not by itself establish which events the adapter prevented.

F Physical Evaluation Protocol

The physical experiments are targeted analogues of the two failure mechanisms studied in simulation. They are intended to test whether the same stage- and risk-localized interaction structure appears on hardware, rather than to constitute a large-scale real-world benchmark.

F.1 Robot and Task Setup

Experiments use a Franka Panda arm with a parallel-jaw gripper and a wrist-mounted RGB camera. Policies output Cartesian end-effector translation and rotation deltas together with a gripper command at 10 Hz. Commands are executed through an operational-space Cartesian impedance controller with limits matched to the simulation action interface. Physical episodes are capped at 200 control steps.

Table 12: **Reported interaction-state diagnostics.**

Module	Metrics
Stage estimator	Regime-wise timestep accuracy
Environment hazard	ROC–AUC; average precision; reliability curve
Robot hazard	ROC–AUC; average precision; reliability curve
Mechanics: continuous	Pearson r for access, progress, and both risk changes
Mechanics: binary	ROC–AUC for attachment stability and snag
Residual adaptation	Event-aligned hazard, gate, and proposed residual norm

The thin-shell analogue uses deformable EVA foam arranged to form a narrow yielding fold pocket. Retrieval requires creating access and extracting the target without tearing the foam or producing excessive structural displacement. The branching analogue uses rubber-coated wires mounted to a fixed base to form a constrained, snag-prone extraction corridor. Retrieval should avoid tangling, structural detachment, and large robot-side contact events.

The environment-risk proxy is computed from visual tracking of occluder displacement and deformation using the wrist camera and an external camera. Robot-side risk uses measured Franka torque/contact signals and controller safety-event flags.

External sensing and calibration. A fixed external camera observes the full interaction workspace while the wrist camera provides a close-range view of the current contact region. Camera intrinsics are calibrated offline, and the external-camera frame is registered to the robot base using a planar fiducial calibration target rigidly attached to the workspace. The calibration is fixed for all trials.

At the beginning of each episode we record a reference configuration of the yielding structure. Environment-side visual risk is computed from tracked occluder points relative to this reference after removing rigid motion of the fixed support. The external view provides global structural displacement, while the wrist view provides local deformation around the active contact region. These signals are normalized using the same conservative-anchor calibration principle as the simulation metrics. Visible tearing or structural detachment is treated as a hard environment violation regardless of the continuous proxy value.

F.2 Trial Protocol and Labels

Each backbone and SRAH variant is evaluated on 24 trials per physical analogue. Variants reuse the same collection of initial target poses and reset configurations to reduce configuration-induced variance.

A trial is a *raw success* when the target reaches the goal region within the episode horizon. It is a *gentle success* when raw success is achieved **without** an environment-side or robot-side violation. Environment violations include visible tearing, structural detachment, excessive displacement, persistent tangling, or other clear stress marks in the yielding structure. Robot violations are defined from contact events, torque/contact thresholds, or controller safety reflexes. Ambiguous physical events are conservatively counted as violations.

Reset and material replacement. Each physical analogue is mounted in a fixed jig with reference marks defining the nominal reset configuration. After every trial, the target and yielding structure are manually returned to this reference configuration and allowed to relax before the next rollout. A reset is accepted only when the tracked structure can be returned to within 5 mm of its reference geometry.

Any element exhibiting visible tearing, detachment, coating failure, or permanent deformation that cannot satisfy this reset criterion is replaced before subsequent trials. Trials that produce such irreversible changes remain in the evaluation set and are labeled as environment violations; they are never repeated on a damaged structure. This procedure prevents later methods from benefiting from or being penalized by damage accumulated during earlier evaluations.

Table 13: **Per-backbone physical component ablations.** Counts are out of 24 trials for each backbone, analogue, and variant.

Backbone	Variant	Thin-shell analogue		Branching analogue	
		Raw/Gentle ↑	E/R viol. ↓	Raw/Gentle ↑	E/R viol. ↓
Diffusion Policy	Base	13/9	4/3	13/7	4/5
	+ stage	13/10	3/3	13/8	4/4
	+ stage + budget	13/11	2/2	13/9	3/3
	+ stage + budget + hazard	13/11	2/1	14/10	2/3
	Full SRAH	**14/12**	**2/1**	**14/11**	**2/3**
ACT-style	Base	14/11	3/2	14/8	3/4
	+ stage	14/12	3/2	14/9	3/3
	+ stage + budget	14/13	1/1	14/11	2/2
	+ stage + budget + hazard	14/13	1/1	14/12	2/2
	Full SRAH	**15/14**	**1/1**	**15/13**	**1/2**
OpenVLA-ft	Base	16/13	3/2	15/10	3/3
	+ stage	16/14	2/2	15/11	2/3
	+ stage + budget	16/15	2/1	15/13	1/2
	+ stage + budget + hazard	16/15	2/1	15/14	1/1
	Full SRAH	**16/16**	**1/0**	**15/14**	**1/1**

Physical training provenance. The physical study is not a zero-shot sim-to-real evaluation. Each backbone is initialized from its corresponding simulation-trained checkpoint and adapted to the physical analogue using the same held-out physical demonstration set and its native training objective. The resulting backbone is then frozen. SRAH is trained separately for each frozen backbone on a disjoint physical training split using the same supervision structure as in simulation: gentle reference trajectories supervise residual imitation, while gentle, unsafe, and failed rollouts supervise stage, mechanics-response, and short-horizon hazard prediction. Evaluation target poses and reset configurations are held out from both backbone adaptation and SRAH training, and no evaluation trajectory is used for gradient updates or model selection. Thus each Base versus Base+SRAH comparison uses exactly the same frozen backbone; the only difference is the learned interaction-state adapter. All Base and Base+SRAH physical comparisons share the identical frozen backbone checkpoint, controller, sensing stack, and evaluation configurations.

F.3 Per-Backbone Physical Results

Table 13 reports physical component ablations separately for each backbone. Across both analogues and all three backbones, Full SRAH adds 22 gentle completions but only four raw completions relative to Base. This supports improved retrieval acceptability; the aggregate counts alone do not identify which individual unsafe trials became gentle successes.

G Benchmark Robustness and Baseline Fairness

This section tests whether GRAM's principal conclusions depend on a particular baseline implementation, simulator discretization, or material/contact setting.

G.1 Baseline Interfaces and Fairness

All learned methods use the same benchmark train/validation partitions and are evaluated under identical physics, sensing, task definitions, action limits, and budgeted success criteria. Learned methods are trained on the same demonstration mixture unless the method itself requires a different learning objective. All methods ultimately execute through the shared incremental Cartesian end-effector interface plus a gripper command.

Table 14 summarizes the principal methodological differences retained by each baseline.

For learned baselines, model selection uses fixed validation subsets rather than test performance. The benchmark preserves native policy structure where it materially defines the method; for example,

Table 14: **Baseline objectives and interfaces.** The benchmark standardizes the executed Cartesian action interface while preserving each method's native internal parameterization where possible.

Method	Internal action representation	Training objective	Budget-aware?	Default inputs / role
Scripted+Imp	Hand-coded state machine + impedance	None	N/A	RGB-D + proprioception; heuristic reference
Diffusion Policy	Continuous SE(3) increments from diffusion model	Behavioral cloning / diffusion likelihood	No	RGB-D + proprioception + wrench estimate
PerAct-style	Discrete voxel action + rotation bins, decoded to SE(3)	Cross-entropy over discretized actions	No	RGB-D voxel workspace + proprioception
ACT-style	Chunked continuous SE(3) sequence	Sequence imitation	No	RGB-D + proprioception + wrench estimate
OpenVLA-ft	Continuous SE(3) increments from adapter head	Supervised fine-tuning	No	RGB or RGB-D, with optional language
COptiDICE	Continuous SE(3) increments	Constrained offline RL	Yes	Same observation suite as BC baselines
Planner+Compliance	Waypoint planner + compliant execution	None	Test-time	Depth-derived occluder estimate + proprioception
Oracle-MPPI	MPPI-sampled SE(3) increments	Online budget-penalized MPC	Yes	Privileged target and occluder state; upper-bound reference

Table 15: **Full GRAM composite-suite gentle success (%).** Each entry is the mean success rate; the main paper focuses on Cross-family and Tightened-budget because these regimes directly test the benchmark findings that motivate SRAH.

	Thin-shell folds					Branching corridors				
Method	Match	Within	Cross	Tight	Capture	Match	Within	Cross	Tight	Capture
Scripted+Imp	42.1	34.0	21.5	18.2	39.4	54.4	43.1	28.6	25.0	50.9
Planner+Compliance	55.2	48.3	30.1	33.7	52.4	70.0	60.2	40.4	45.6	66.3
Diffusion Policy	67.9	55.6	31.7	27.9	63.1	74.1	62.0	35.4	30.7	69.0
PerAct-style	64.8	52.2	29.8	34.5	60.4	78.3	66.4	42.1	38.2	72.1
ACT-style	70.3	58.7	35.9	45.1	66.9	80.2	67.9	44.9	49.8	75.8
OpenVLA-ft	65.8	56.9	43.8	52.3	63.9	71.9	60.7	52.0	54.3	70.2
COptiDICE	58.4	49.0	33.2	44.0	55.6	63.5	53.4	36.6	47.2	61.0
Oracle-MPPI	82.4	74.1	55.6	60.8	78.2	88.0	80.1	65.0	70.2	84.3

PerAct-style retains its voxel action representation before decoding to the common executed SE(3) interface.

G.2 Full Composite-Suite Success Overview

Table 15 reports the full composite-suite gentle-success results underlying the benchmark analyses. We include the table here primarily to show that the main-paper Cross-family and Tightened-budget comparisons were not selected from an isolated favorable regime.

The complete table reinforces two observations used in the main paper. Morphology shift produces substantially larger degradation than Capture shift for most learned methods, and Tightened-budget evaluation can change relative method behavior because raw task competence and risk-sensitive interaction are not equivalent.

G.3 Simulator and Modeling Robustness

We test the benchmark under perturbations of physics timestep, occluder discretization, contact/material parameters, and solver execution. Unless otherwise noted, these studies use fixed episode sets from the Tightened-budget composite suite.

Physics timestep. The physics timestep is varied over

$$\Delta t \in \left\{ \frac{1}{1200}, \frac{1}{800}, \frac{1}{600}, \frac{1}{400} \right\} \mathrm{s}, \tag{33}$$

with controller gains adjusted to preserve comparable closed-loop bandwidth in physical time. The default is $1/800$ s. Across the representative baseline set, performance changes smoothly and degrades only modestly at the coarsest timestep, where contact resolution is reduced.

Table 16: **Ranking stability under contact and material parameter sweeps.** Entries report mean / minimum / maximum Kendall τ against the default method ordering.

Family	Sweep	Success	EnvDist violation	RobotCost violation
Thin-shell	Shell damping $\times\{0.7, 1.0, 1.3\}$	0.90 / 0.84 / 0.94	0.92 / 0.86 / 0.96	0.93 / 0.88 / 0.97
Thin-shell	Contact stiffness $\times\{0.8, 1.0, 1.2\}$	0.89 / 0.82 / 0.93	0.91 / 0.85 / 0.95	0.92 / 0.87 / 0.96
Thin-shell	Friction $\mu \in \{0.6, 0.8, 1.0\}$	0.88 / 0.80 / 0.92	0.90 / 0.83 / 0.94	0.91 / 0.85 / 0.95
Branching	Branch stiffness $\times\{0.8, 1.0, 1.2\}$	0.89 / 0.83 / 0.93	0.91 / 0.85 / 0.95	0.92 / 0.86 / 0.96
Branching	Friction $\mu \in \{0.6, 0.8, 1.0\}$	0.87 / 0.79 / 0.92	0.90 / 0.82 / 0.94	0.90 / 0.83 / 0.95
Branching	Contact damping $\times\{0.8, 1.0, 1.2\}$	0.90 / 0.84 / 0.94	0.92 / 0.86 / 0.96	0.93 / 0.88 / 0.97

Occluder discretization. Thin-shell surfaces are evaluated at approximately 8k, 15k (default), and 30k vertices while preserving the underlying procedural geometry. Branching structures are evaluated with skeleton segment lengths of 2.5 cm, 1.5 cm (default), and 1.0 cm. Across these changes, success and violation rates remain close to the default setting, indicating that the principal results do not depend on a particular mesh or skeleton resolution.

Contact and material parameters. Table 16 reports method-ranking stability under material and contact sweeps. Rankings remain strongly correlated with the default configuration across both success and violation metrics.

Solver randomization. We additionally perturb contact-solver ordering and warm-start state while keeping physical parameters fixed. Across the tested methods, micro-penetration occupies approximately 1.0–1.8% of simulated steps and solver randomization reduces success by only 0.6–1.6 percentage points. The effect is small relative to the benchmark's cross-method and cross-regime gaps, and no method obtains low disturbance through anomalously high penetration.

Together, these checks indicate that GRAM's qualitative conclusions are stable to the principal simulator, discretization, and contact-model choices tested here rather than being artifacts of a single numerical configuration.

H Additional Benchmark Analyses

This section preserves two secondary GRAM analyses that provide additional interpretation of the benchmark complementary the main SRAH claims. The first asks whether failures are primarily caused by missing information or by interaction strategy. The second asks which structural properties of the yielding occluder are predictive of retrieval difficulty.

H.1 Observation Ladder: Information Versus Strategy

We evaluate an ACT-style policy on hard-tier tasks while progressively adding sensing modalities. Table 17 reports gentle success together with normalized environment disturbance. Increasing observability consistently improves performance and reduces disturbance, but the magnitude of the gain depends on the interaction regime.

Thin-shell fold-pocket and dynamic tasks benefit strongly from depth, proprioception, wrist force/torque, and tactile sensing, consistent with severe self-occlusion making state estimation a substantial bottleneck. Branching tasks also improve with richer sensing, but the gains are smaller and substantial failures remain even with strong contact information. This suggests that snag-prone extraction retains a larger strategy and recovery component rather than being explained solely by missing observations.

The oracle-target-pose condition is diagnostic rather than a deployable policy input. Its remaining gap from perfect success indicates that complete target localization does not remove the contact-planning and recovery difficulty of these tasks.

Table 17: **Observation ladder on hard-tier retrieval tasks.** Each cell reports gentle success (%) / normalized EnvDist for an ACT-style policy. Additional sensing improves both task completion and disturbance control, with larger information gains in deeply occluded thin-shell tasks.

	Thin-shell folds		Branching corridors	
Observation suite	Fold-Pocket	Dynamic	Snag-Aware	Corridor
RGB	11.8±1.6 / 0.88	17.6±1.8 / 0.86	27.9±2.0 / 0.82	34.6±2.1 / 0.80
RGB-D	23.9±1.9 / 0.82	30.4±2.1 / 0.80	40.3±2.2 / 0.77	47.8±2.3 / 0.75
RGB-D + proprioception	35.7±2.1 / 0.76	41.9±2.2 / 0.74	45.8±2.2 / 0.74	54.1±2.3 / 0.72
+ wrist F/T	42.1±2.2 / 0.72	47.8±2.3 / 0.70	55.2±2.3 / 0.68	59.8±2.3 / 0.67
+ tactile	55.0±2.3 / 0.64	58.4±2.3 / 0.63	61.7±2.3 / 0.64	65.2±2.3 / 0.63
Oracle target pose	78.1±2.0 / 0.68	73.8±2.1 / 0.67	80.4±1.9 / 0.66	82.2±1.8 / 0.65

Table 18: **Task-conditioned local structure is more predictive of retrieval outcomes than coarse global geometry.** Entries are ROC–AUC for predicting success and the characteristic failure outcome in each family, averaged over four seeds. Parentheses denote $100\times$ standard deviation.

	Thin-shell folds		Branching corridors	
Predictor	Success	Budget violation	Success	Extraction failure
Target visibility	0.79(2)	0.64(3)	0.68(2)	0.57(4)
Best-grasp clearance	0.83(2)	0.71(4)	0.77(2)	0.69(3)
Local curvature spectrum	0.86(2)	0.82(2)	0.58(3)	0.54(3)
Pocket-mouth width / fold aperture	0.88(2)	0.85(2)	0.55(3)	0.52(3)
Local branch density	0.56(3)	0.53(1)	0.84(2)	0.81(1)
Corridor-width distribution	0.60(3)	0.56(3)	0.86(2)	0.84(2)
Snag-risk index	0.58(3)	0.55(3)	0.88(2)	0.90(2)
Global genus / hole count	0.54(3)	0.52(2)	0.55(3)	0.54(3)
Total surface area / branch length	0.56(3)	0.53(3)	0.57(3)	0.56(2)
Global connectivity / component count	0.55(3)	0.54(5)	0.58(3)	0.57(3)

H.2 Local Structure Predicts Retrieval Difficulty

We next ask whether retrieval outcomes are explained by coarse scene complexity or by geometry local to the target and plausible interaction corridor. Table 18 reports the predictive value of individual structural descriptors using ROC–AUC.

For thin-shell retrieval, pocket-mouth width, local curvature, and best-grasp clearance are strongly predictive of both success and budget violation. For branching retrieval, corridor width, local branch density, and the snag-risk index are most predictive of success and extraction failure. In contrast, global descriptors such as total surface area, branch length, topology, or component count remain near chance once the benchmark difficulty distribution is controlled.

These results support the proposed interpretation of the benchmark: retrieval difficulty is governed primarily by the local mechanics and geometry of the interaction region rather than by coarse visual or geometric scene complexity.

Taken together, these two analyses refine the main benchmark diagnosis. Thin-shell retrieval contains a substantial information-access component, while branching retrieval retains a stronger contact-strategy and recovery component. Across both families, the structural variables that matter most are local to the interaction region, consistent with representing current mechanics response rather than global morphology identity.