FROM CONTACT TO ACTION: FOLD-CONTACT AFFORDANCES FOR CLOTH UNFOLDING

Anonymous authors
Paper under double-blind review

ABSTRACT

Manipulation depends on physical relationships that visible shape alone can leave ambiguous. Cloth unfolding makes this problem concrete: the same overhead silhouette can conceal different arrangements of folds and support. We study how complementary sensing can expose support cues and make them useful for action. A transparent-table dual-arm setup pairs a top segmentation mask with a bottom contact mask. We derive fold residual, contact overlap, fold-contact frontier, and silhouette-edge maps for an attention encoder and a residual action prior. An online linear, contextual-bandit-style scorer adapts the prior's target ranking from reward, self-supervised unfolding progress, or their mixture. In the long-horizon comparison, raw dual-view pixels improve flatness over a top-only policy; affordance channels reduce folded area by 82.2% relative to raw dual masks in a fast screen; and the reward-bandit bottom-view scorer performs best in the 21-run residual-prior ablation, with within-panel geometry score 0.840 ± 0.047. Additional randomized-state and physical evaluations extend these comparisons; calibration and illumination studies characterize the sensing requirements. These findings suggest a design principle for cloth unfolding: organize perception around support relationships and connect that representation to target selection through interaction feedback.

1 INTRODUCTION

Robotic manipulation requires choosing actions from incomplete visual evidence. A central difficulty is that physical states with similar visible shapes can require different actions: contact and support influence what can move and where an action should be applied. This motivates a question of observability for action: which hidden physical relationships should sensing expose, and how can a policy use them to choose where to act?

Cloth unfolding makes this question concrete. A nearly flat garment, a stack of folds, and a lifted flap can share the same overhead silhouette, yet differ in how the cloth is supported. A useful dual-arm pull depends on which region is grounded and which can be drawn away. This ambiguity also affects the apparent progress of a reinforcement learning (RL) policy: enlarging the silhouette can leave folded layers unresolved. Continuous measures of folds, contact, and flatness therefore complement coverage when assessing unfolding progress.

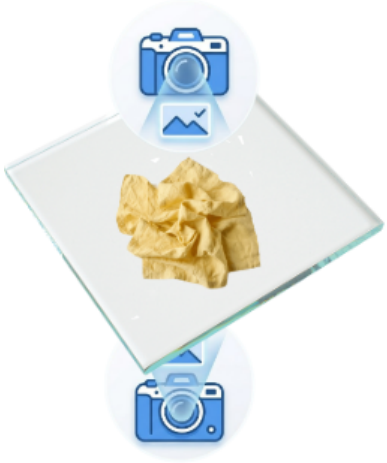


Figure 1: Top silhouette and bottom contact.

We investigate this missing support information in a transparent-table setting that pairs a top-view cloth silhouette with a bottom-view contact mask, or the corresponding simulated masks. The bottom view identifies occupied cells near the table surface, supplying partial information about support without reconstructing the full cloth configuration. We ex-

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.

press the relationship between the masks as fold-contact affordances: residual folded area, grounded contact, the frontier between folded and supported regions, and the outer silhouette edge.

We connect this support representation to control in two ways. A contact-affordance attention encoder gives the policy explicit access to the cues, while the fold-contact frontier guides candidate grasp locations for a residual action prior. The prior directs exploration toward geometrically meaningful targets, and the SAC actor learns continuous corrections. The representation thus gives support information a role in both state encoding and action proposals. The maps provide a parameter-free geometric inductive bias; SAC still learns the action policy end to end. In the representation screen, these structured cues yield larger improvements in fold and contact quality than raw dual-view pixels.

A geometrically plausible target does not guarantee unfolding progress. We use interaction feedback to adapt the ranking criterion, which remains fixed under hand-tuned frontier weights. A lightweight online linear scorer updates target preferences in a contextual-bandit-style manner from reward, self-supervised unfolding progress, or a mixture of the two. This connects observing support to learning where to act, while leaving continuous control to the residual RL policy.

This paper makes three contributions that connect support sensing to unfolding:

1. **Representing support.** A fold-contact observation model converts top and bottom masks into residual fold, contact, frontier, and edge maps, exposing support cues absent from the overhead silhouette.
2. **Guiding action.** A bottom-view residual action prior selects targets using the fold-contact frontier while SAC learns continuous corrections for dual-arm motion.
3. **Adapting target selection.** A learned affordance scorer replaces fixed frontier weights with an online bandit-style update. The best variant learns from environment reward, raises the within-ablation geometry score from 0.592 to 0.840 over a fixed heuristic residual prior, and reduces folded area by 65.4% while keeping area nearly unchanged.

2 Related Work

Cloth manipulation combines high-dimensional state, self-contact, and partial observability, making task structure important for perception and control (Sanchez et al., 2018; Longhini et al., 2025). Geometric manipulation sequences and learning from demonstrations or scripted supervision provide complementary approaches to this challenge (Maitin-Shepard et al., 2010; Cusumano-Towner et al., 2011; Jia et al., 2019; Matas et al., 2018; Seita et al., 2020). Dynamic flinging, flow-based bimanual control, and learned folding primitives illustrate the value of visual and action priors (Ha & Song, 2022; Weng et al., 2022; Avigal et al., 2022). Our focus is how support information from complementary sensing can inform both a learned representation and manipulation targets.

Partial-state methods infer cloth connectivity or reconstruct occluded geometry (Lin et al., 2022; Chi & Song, 2021; Huang et al., 2022). We instead expose a partial support relation through projected masks. Affordance learning and spatial action representations motivate making that relation explicit (Zeng et al., 2018; Florence et al., 2018; Zeng et al., 2021). SAC supplies the continuous controller (Haarnoja et al., 2018); a lightweight target scorer draws on online feedback and exploration ideas from contextual bandits (Li et al., 2010; Auer et al., 2002), without asserting a regret guarantee. Appendix B gives the extended discussion.

3 Method

Figure 2 gives an overview of the pipeline.

Task. The simulated task uses two end effectors and a crumpled short-sleeve garment on a table. At each control step the action is

$$a_t = [\Delta x^L, \Delta y^L, \Delta z^L, \Delta \theta^L, \Delta g^L, \Delta x^R, \Delta y^R, \Delta z^R, \Delta \theta^R, \Delta g^R].$$

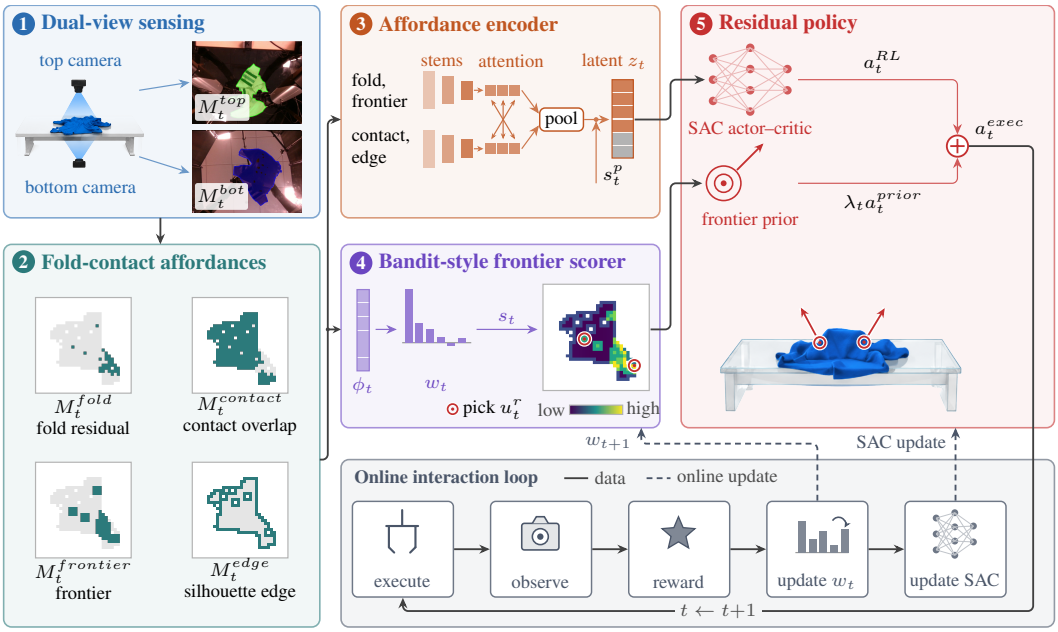


Figure 2: From contact to action. **(1)** Dual-view sensing: the top camera records the silhouette mask M_t^{top} and the bottom camera, looking up through the transparent table, records the contact mask M_t^{bot}. **(2)** Fold-contact affordances: parameter-free morphology converts the two masks into the fold residual M_t^{fold}, contact overlap $M_t^{contact}$, fold-contact frontier $M_t^{frontier}$, and silhouette edge M_t^{edge}. **(3)** Affordance encoder: separate stems for the fold/frontier and contact/edge channels exchange tokens through bidirectional attention, are pooled, and are concatenated with proprioception s_t^p into a compact latent z_t that is trained end-to-end with the actor and twin critics. **(4)** Bandit-style frontier scorer: candidate cells are ranked by $s_t(i,j) = w_t^\top \phi_t(i,j)$ and the best cell on each arm's side, u_t^r for $r \in \{L, R\}$ (circles), becomes that arm's pick target. **(5)** Residual policy: the SAC action a_t^{RL} and the scaled per-arm frontier prior $\lambda_t a_t^{prior}$ are summed and clipped into the executed dual-arm action a_t^{exec} (circles: pick targets; red arrows: outward pull). The lower loop summarizes online interaction: execute the action, observe the next cloth state, compute reward and progress, update the frontier scorer ($w_t \rightarrow w_{t+1}$), and update the SAC actor–critic from experience; solid arrows carry data and dashed arrows are online updates.

The proprioceptive state contains end-effector poses, target poses, gripper openings, and joint positions. The reward is grounded: area, spread, and flatness gains are gated by cloth height above the table, with penalties for action magnitude, grasp stalls, and pulling without cloth progress.

Dual-view masks. Let M_t^{top} be the top-view cloth silhouette over a fixed table-centered grid. In simulation, the bottom mask is computed from cloth particles and table height; it is therefore privileged relative to a real upward-facing camera. The mask is active where the lowest cloth surface in an occupied cell lies near the table:

$$M_t^{bot}(i,j) = \mathbb{1}\left[M_t^{top}(i,j) = 1 \wedge z_{i,j}^{min} \leq z_{table} + \epsilon_{bot}\right].$$

On the real robot, M_t^{top} is an RGB segmentation and M_t^{bot} applies the same height test to the cloth-underside height estimated from bottom-view depth (Appendix G); in both domains the masks are near-table support proxies rather than exact physical-contact ground truth. The top-only baseline observes $[s_t^p, M_t^{top}]$, while the raw dual-view baseline observes $[s_t^p, M_t^{top}, M_t^{bot}]$. We add mask dropout and random flips during training, and the occlusion stress test masks the policy observation; unmodeled real effects such as calibration drift, reflections, and segmentation bias remain limitations of this simulation model. Physical registration and illumination tests are reported separately in Appendices S and T.

Fold-contact affordance channels. The affordance observation makes support cues explicit by separating the top-visible residual from table contact:

$$M_t^{fold} = \max(M_t^{top} - M_t^{bot}, 0), \qquad M_t^{contact} = M_t^{top} M_t^{bot},$$
$$M_t^{frontier} = \mathrm{dilate}(M_t^{fold})\,\mathrm{dilate}(M_t^{contact}) M_t^{top}, \qquad M_t^{edge} = M_t^{top} - \mathrm{erode}(M_t^{top}).$$

The implementation uses one iteration of a 3×3 square binary kernel for each dilation/erosion operation, a 32×32 mask grid, and $\epsilon_{bot} = 3.5$ cm unless otherwise stated. The policy receives $[s_t^p, M_t^{top}, M_t^{bot}, M_t^{fold}, M_t^{contact}, M_t^{frontier}, M_t^{edge}]$. A contact-affordance attention encoder applies separate visual stems to fold/frontier and contact/edge channels, exchanges tokens with bidirectional attention, pools them, and concatenates the result with proprioception before the actor and critic. Appendix F gives the exact channel widths, tokenization, and checkpoint-selection protocol. The mask operations introduce no learned parameters; the visual encoder, actor, and critic remain learned components. Binary near-table occupancy identifies neither layer count nor all cloth–cloth contacts, and M_t^{fold} is a support residual rather than a detector of every fold, so these channels encode support cues rather than a complete reconstruction of the hidden cloth state. Appendix W summarizes the notation.

Residual bottom-view action prior. The fold-contact frontier turns the support representation into an action prior. For each arm, candidate cells are restricted to the corresponding side of the garment. A cell feature vector

$$\phi_t(i,j) = [M^{frontier}, M^{edge}, M^{fold}, M^{top}, M^{contact}, d^{out}]$$

is scored, the best cell becomes a pick target, and the prior moves the gripper toward the target, closes near it, and pulls outward from the garment center. The executed action is residual:

$$a_t^{exec} = \mathrm{clip}\Big(a_t^{RL} + \lambda_t a_t^{prior}, -1, 1\Big).$$

Learning the scorer. The fixed prior uses hand-chosen weights over ϕ. We replace them with a linear, contextual-bandit-style scorer $s_t(i,j) = w_t^\top \phi_t(i,j)$. For each arm $r \in \{L, R\}$, the candidate set $\mathcal{C}_t^r$ contains top-mask cells on that arm's side of the table-centered ROI with finite score; if all learned scores are non-positive, the implementation falls back to a top-mask plus outward-distance score on the same side. The selected cell $u_t^r = (i_t^r, j_t^r)$ is greedy except for ϵ-exploration, which samples uniformly from the top quartile of finite-scored candidates on that side. At each policy step we form bounded feedback $p_t \in [-1, 1]$ from either grounded geometry progress g_t, reward $\tanh(R_t/2)$, or their mixture. With EMA baseline $\bar{a}_t$, clipped advantage $A_t = \mathrm{clip}(p_t - \bar{a}_t, -1, 1)$, and centered feature delta $\Delta\phi_t$ between selected frontier cells and their candidate-set mean, the weights are updated by

$$w_{t+1} = \Pi_{[l,u]}\big[w_t + \eta A_t\, \Delta\phi_t + \eta\rho(w_0 - w_t)\big],$$

where $\Pi_{[l,u]}$ clips each coordinate. The bounded feedback, centered features, decay toward the heuristic w_0, and projection keep each update bounded; we use η=0.20, ρ=0.01, β=0.90, and ϵ=0.10 in the reward-bandit runs. This is best viewed as a projected advantage update for target ranking, not as a bandit algorithm with a regret guarantee. Appendix P.1 gives the exact pseudocode, stability argument, and UCB-style uncertainty variant.

4 Experiments

The evaluation asks how additional support information becomes useful for control: through complementary sensing, explicit representation, action proposals, and learned target ranking.

Metrics. To distinguish visible coverage from physical unfolding, we report continuous progress metrics alongside evaluation return. Binary success depends on the garment, initial crumple, and chosen coverage or flatness threshold; the geometry measures identify which aspect of unfolding changes: final cloth area A for coverage, folded mask area F for residual folds, contact ratio C for grounded support, and height standard deviation Z for flatness. For screening we use

$$S = 0.35A + 0.25C + 0.25(1 - F) + 0.15(1 - Z),$$

where each component is min-max normalized only within the variants in the same table or figure; higher area/contact and lower fold/z-std are better. We use S only as a within-panel ranking index; cross-block comparisons should use the absolute return, area, fold, contact, and z-std columns.

Evaluation scope. The main learning and scorer studies use a shared fixed initial state and their stated training horizons. A separate randomized campaign and quantitative physical evaluation test broader configurations; their results are reported separately below. Extended diagnostics appear in Appendix A.

4.1 From Sensing to Support Representation

The one-million-step comparison tests the information added by bottom view. Raw dual-view cross-attention only slightly changes return and area relative to the top-mask baseline, but it lowers height standard deviation from 0.0271 to 0.0218, a 19.8% reduction. This comparison shows a flatness benefit that return and coverage alone would obscure, as summarized in Figure 3.

The fast screen then examines how the added information is represented. Raw dual masks obtain high area but leave substantial folds: folded area 0.0728, contact ratio 0.659, and z-std 0.0467. The affordance observation lowers folded area to 0.0130 and z-std to 0.0231, while increasing contact ratio to 0.931. This representation comparison favors making support/contact structure explicit: the clearest improvements concern residual folds and grounded contact.

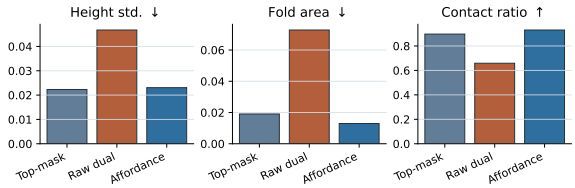


Figure 3: From complementary sensing to a support representation. Raw dual view improves some flatness diagnostics over top-only observation; the fold-contact affordance decomposition sharply reduces folded area and improves grounded contact in the fast screen. The top-mask fold/contact bars are filled from the current-metric top-only evaluation because the earliest one-million-step top run did not log these bottom-derived diagnostics.

4.2 Representation and Sensing under Randomized Configurations

A separate campaign uses 3 seeds, 14 randomized training states, and 17 held-out physical drop-crumples. Table 1 and Figure 4 (top) report its folded area and contact results. The dual-affordance policy gives $4.5\times$ less fold and $+.18$ contact relative to top-only SAC. Under matched sensing, action space, reward, network, and training budget, replacing raw dual-view pixels with affordances changes F/C from $.0466/.766$ to $.0066/.968$: $7.1\times$ less fold and $+.202$ contact. This comparison supports the value of the geometric representation within an end-to-end learned policy.

Adding bottom sensing to the affordance representation reduces folded area by a further 25% and raises contact by $+.011$ relative to the same representation computed from top depth alone. These results belong to the randomized campaign and are not pooled with the earlier screens. Appendix Q gives the full decomposition and reporting conventions.

Table 1: Randomized campaign on 17 held-out states. Reported uncertainties are standard errors.

Method	Folded area $F \downarrow$	Contact $C \uparrow$
Dual affordance (ours)	$.0066 \pm .001$	$.968 \pm .007$
Top-depth affordance	$.0088$	$.957$
Top-only SAC	$.0297 \pm .007$	$.790 \pm .060$
Fling-style	$.0357$	$.688$
Raw dual SAC	$.0466$	$.766$
VLM pull	$.0526$	$.558$

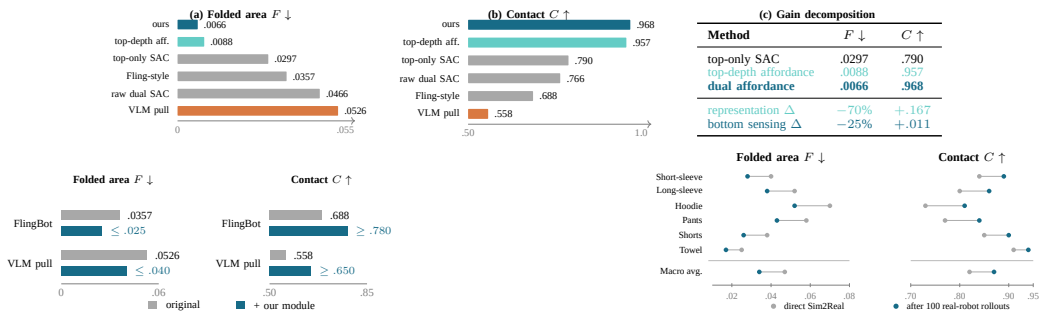


Figure 4: Validation plots. Top: randomized-campaign folded area, contact, and gain decomposition (Table 1). Bottom left: policy-module comparisons under matched actuator, budget, and tests (Table 2). Bottom right: direct sim-to-real transfer and performance after real-robot fine-tuning, with category-level points and the macro average (Appendix R).

4.3 From Support Cues to Action Proposals

A no-learning 80-step action-prior probe shows that bottom-view frontiers are actionable: the affordance prior increases area by 0.0469 versus 0.0072 for zero action and 0.0052 for random dual actions. However, the pure prior can slightly hurt contact and fold quality. Longer residual-policy training confirms the trade-off: a 12-hour residual-prior run improves the geometry score over affordance observation alone (0.650 versus 0.350), mainly through lower fold, higher contact, and lower z-std, but sacrifices some area and return. These trade-offs motivate learning how to rank frontier targets from interaction, so that geometric proposals can adapt to their unfolding outcomes.

Compatibility with other manipulation policies. The support-based module also improves a fling-based policy and VLM-guided pulling under matched actuator, budget, and tests (Table 2; Figure 4, bottom left). The fling-based policy is a FlingBot-style (Ha & Song, 2022) baseline whose small fully convolutional value network scores pick pairs on the 32×32 top-mask grid; the VLM-guided baseline asks Qwen2.5-VL-7B-Instruct (Bai et al., 2025) for a pick pair from a rendered top-view height image. Both execute the pair with the same scripted stretch-and-fling primitive. The module is also complementary to fling initialization: a fling followed by the proposed policy improves F/C from $.0066/.968$ to $.0050/.978$.

Table 2: Compatibility with other policies and with fling initialization. The first two rows add the module to the indicated policy; the final row adds fling initialization before executing the proposed policy.

	Without module		With module / initialization	
Policy	$F \downarrow$	$C \uparrow$	$F \downarrow$	$C \uparrow$
FlingBot + module	.0357	.688	.025	.780
VLM pull + module	.0526	.558	.040	.650
Ours with fling initialization	.0066	.968	.0050	.978

4.4 Learning to Rank Manipulation Targets

Table 3 and Figure 5 summarize the 21-run ablation. The reward-bandit scorer performs best within this ablation, with geometry score 0.840 ± 0.047, return 33.4 ± 1.9, folded area 0.0068 ± 0.0035, contact ratio 0.964 ± 0.018, and z-std 0.0173 ± 0.0013. Compared with the heuristic residual prior, it raises the within-ablation score from 0.592 to 0.840 and reduces folded area by 65.4% while keeping area nearly unchanged.

The adaptive gate is useful for progress-only self-supervision: adding it raises the progress scorer from score 0.706 to 0.823 and stabilizes the three seeds. It does not improve the reward scorer, suggesting that reward already penalizes the late over-pulling behavior that the gate was designed

Table 3: Learned bottom-view scorer ablation. Score uses area/contact positively and fold/z-std negatively; values are mean ± standard error over three seeds.

Variant	Runs	Score	Eval return	Area	Fold ↓	Contact ↑	Z-std ↓
Reward-bandit scorer	3	0.840 ± 0.047	33.4 ± 1.9	0.213 ± 0.004	0.0068 ± 0.0035	0.964 ± 0.018	0.0173 ± 0.0013
Progress scorer + adaptive prior	3	0.823 ± 0.010	30.7 ± 0.7	0.213 ± 0.002	0.0072 ± 0.0002	0.963 ± 0.002	0.0181 ± 0.0004
Hybrid progress/reward scorer	3	0.775 ± 0.119	27.1 ± 4.8	0.203 ± 0.012	0.0054 ± 0.0018	0.968 ± 0.013	0.0209 ± 0.0035
Hybrid scorer + adaptive prior	3	0.730 ± 0.179	30.9 ± 6.5	0.200 ± 0.022	0.0064 ± 0.0019	0.958 ± 0.017	0.0218 ± 0.0048
Reward scorer + adaptive prior	3	0.712 ± 0.110	32.3 ± 3.0	0.213 ± 0.011	0.0121 ± 0.0054	0.934 ± 0.027	0.0226 ± 0.0029
Progress-bandit scorer	3	0.706 ± 0.091	29.2 ± 0.7	0.218 ± 0.006	0.0139 ± 0.0058	0.927 ± 0.030	0.0238 ± 0.0041
Heuristic residual prior	3	0.592 ± 0.159	-4.6 ± 31.8	0.216 ± 0.006	0.0197 ± 0.0101	0.898 ± 0.049	0.0265 ± 0.0059

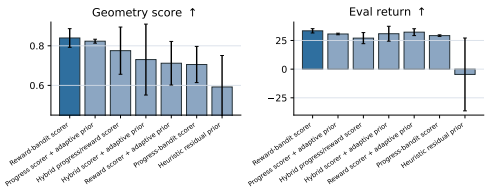


Figure 5: Learned target ranking in the 21-run ablation. The reward-bandit scorer ranks highest in this comparison; adaptive scaling mainly helps the progress-only scorer.

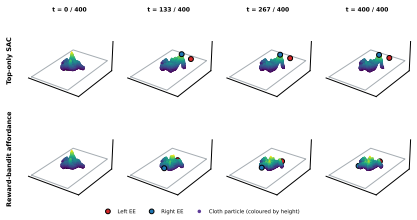


Figure 6: Deterministic dual-arm rollouts of the top-only SAC baseline (top row) and the reward-bandit affordance policy (bottom row) starting from the same crumpled cloth state. The reward-bandit affordance policy reaches a flatter, more grounded configuration with a folded residue roughly $4\times$ smaller than the top-only baseline at the final frame.

to reduce. The hybrid progress/reward scorer gives strong fold/contact in two seeds but has larger variance and lower mean area, so it is not the recommended method. Figure 6 shows deterministic rollouts of the top-only baseline and the reward-bandit affordance policy from the same crumpled cloth state; Appendix D discusses the per-frame metrics.

A UCB-style reward scorer with a count-based exploration bonus is statistically tied with reward-bandit on coverage and z-std at 25k steps, with slightly lower folded area in this small comparison (Appendix O).

4.5 Interpreting the Gains and Their Sensitivity

Under arm-like top-camera occlusion, the reward-bandit affordance policy remains stronger than raw dual-view SAC, raising the within-panel geometry score by 0.152 and return by 16.6 because the bottom mask supplies information missing from the occluded top silhouette.

To interpret the flatness gains, we examine the geometry measures together. Across the 21-run scorer ablation, lower folded area ($r = 0.78$) and higher contact ratio ($r = -0.85$) both accompany lower height standard deviation (Figure 19, Appendix P.2), consistent with the proposed support-based mechanism: improved flatness accompanies fewer residual folds and greater grounded contact.

The reliability of the support cues depends on bottom-mask quality (Table 6). False negatives and random flips are mild: FN 5% and flip 2% change folded area by at most $+0.0013$ and contact by at most -0.009 relative to clean input. A one-pixel registration shift is severe: folded area rises from 0.0109 to 0.0291 ($2.7\times$), contact falls from 0.944 to 0.833, and return drops from 32.8 to -409.7; mixed failure behaves similarly. Retraining with the same perturbation family removes the return collapse (-2.4 under shift, 21.9 under mixed) but does not restore support geometry: contact

Table 4: Quantitative physical evaluation. Coverage gain is final minus initial coverage; means, standard deviations, and ranges follow the evaluation protocol.

Real-world metric	Result
Bottom-contact mask IoU	.91
Precision / recall	.93 / .93
Frontier distance (median / 95th pct.)	1 cm / 2–3 cm
Coverage gain: random	$+.02 \pm .04$
Coverage gain: top-only	$+.07 \pm .05$
Coverage gain: raw dual-view	$+.03$–$+.06$
Coverage gain: dual-view	$+.13 \pm .05$
Dual-view vs. top-only (paired)	$+.06$, 95% CI $[+.02, +.10]$
Dual-view action reduction vs. top-only	21%

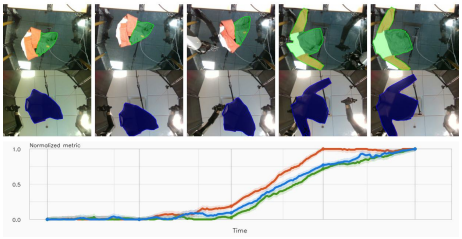


Figure 7: Real-world trial. Top: keyframes segmented by SAM 3 (Carion et al., 2025). Bottom: normalized cloth height decrease, upward lifting, and outward pulling alignment, which increase as the garment unfolds.

stays below the clean-trained policy under every perturbation, and on clean input the retrained policy trades area and return for slightly lower fold and higher contact. Augmentation thus prevents penalty accumulation rather than recovering support-cue quality. Table 6 and its protocol are in Appendices A.1 and N; Appendices K and M report the occlusion and mask-sensitivity results.

Physical calibration targets this registration failure directly: a double-sided board reduces measured top–bottom registration offset and restores mask agreement. The associated folded-area degradation falls from 5–10$\times$ to $\leq 1.1\times$ relative to clean input (Appendix S). Separate illumination tests show stable segmentation under moderate brightness changes and degradation under very low, side, or overexposed lighting (Appendix T).

4.6 Real-World Evaluation

The physical setup (Appendix C, Figure 8) is evaluated with 11 trials per garment–method pair; coverage gain is final minus initial coverage, summarized by mean $\pm$ standard deviation. Dual-view coverage gain is $+.13 \pm .05$, versus $+.07 \pm .05$ for top-only; the paired difference is $+.06$, with 95% CI $[+.02, +.10]$. Dual view also reduces actions by 21% relative to top-only (Table 4). Because F is near zero for a garment folded flat on the glass, neither F nor coverage gain certifies unfolding on its own. Table 5 therefore reports the physical evaluation by initial state (final normalized coverage NC and unfolding success, where success requires the state's NC standard *and* an independent fold check, over 20 independent resets per state) together with the contact-perception and registration quality that the action prior relies on over 60 test configurations. Random crumples score lower because their difficulty varies more across resets; the spread of final NC shows the effect is not carried by a few favorable trials (protocol in Appendix V).

Protocol and efficiency. The per-state trials of Table 5 start from a flat double fold, a settled partial overhang, a random crumple, or the flat-folded state, each after an independent reset (Appendix V), and timing runs from the first robot action until the success criterion is met. A trial fails if the criterion is not met within 720 s, or if at least 50% of the garment area is dragged out of the top camera's view. An action is one executed pick-and-pull by either arm, and no separate cap on the number of actions is imposed beyond the timeout. The current implementation still needs many such interactions per trial, and we make no claim of execution speed over dynamic unfolding methods; the 21% action reduction relative to top-only (Table 4) indicates that the support-aware representation improves action selection and unfolding outcome. The module is also complementary to existing policies: it improves the F/C metrics of the tested FlingBot-style and VLM baselines (Table 2), although better support metrics alone do not certify complete unfolding.

Unfolding from flat-folded initial states. A garment folded flat with its layers on the glass already has a low support residual f_f and mostly near-table contact while it remains folded, so neither quantity can signal completion. We therefore start 3 garments held out from parameter tuning from this state, with 20 independent resets per garment and method under the success criterion of Ta-

Table 5: Physical evaluation by initial state and perception quality. (a) Final normalized coverage NC, success rate, support residual f_f, and NC spread over 20 independent resets per state. (b) Near-table (contact) mask quality, depth validity, height error, and top–bottom registration error in 2.5 cm policy-grid cells over 60 test configurations. (c) Flat-folded state: 3 garments held out from parameter tuning, 20 independent resets per garment and method; final NC not recorded for top-depth.

(a) Unfolding outcome by initial state

Initial state	Final NC	Success
Flat double fold	0.90	80%
Partial overhang (settled)	0.90	80%
Random crumple	0.85	70%
Final f_f (auxiliary)	0.05	
Std. of final NC across trials	0.08	
Independent trials per state	20	

(b) Near-table perception and registration

Metric	Value
Contact-region IoU	0.91
Precision / recall	0.93 / 0.93
False near-table rate (cloth above ϵ_{bot})	10%
Valid-depth fraction inside the garment	90%
Height error w.r.t. table plane, mean / 95th pct.	10 / 20 mm
Registration error, median / 95th pct.	0.05 / 1 cell
Independent test configurations	60

(c) Flat-folded state, 3 held-out garments

Method	Success	Final NC
Full method (dual affordance)	**51/60 (85%)**	**0.91**
Top-depth affordance	36/60 (60%)	–
Edge-only	39/60 (65%)	0.84

ble 5, and compare the full method with two controls: top-depth affordance, which computes the same channels from top depth alone (Section 4.2), and edge-only, which restricts pick targets to the silhouette-edge channel M_t^{edge} without the fold-contact frontier. The full method succeeds in 51/60 trials (85%) versus 36/60 (60%) for top-depth affordance and 39/60 (65%) for edge-only, gains of 25 and 20 percentage points, and its final NC is 0.91 versus 0.84 for edge-only (Table 5c). Despite the low initial f_f and high initial contact, the full method still unfolds the garment in most trials, so both the bottom view and frontier targeting remain useful in this state; resets are independent rather than paired across methods (Appendix V).

Appendices U and R report garment areas and transfer results (Figure 4, bottom right); Figure 7 shows a representative trial, and full records appear in Appendix E.

5 Limitations & Conclusion

The experiments connect bottom-view sensing to action through an explicit support representation. Fold-contact affordances improve fold, contact, and flatness diagnostics over raw dual masks in the representation screen. The same frontier supplies targets for a residual action prior, and reward-trained ranking performs best in the main scorer ablation. Here the value of complementary sensing lies in exposing support cues and giving them a direct role in action selection.

The representation has a specific observability limit: all tested methods fail on 3–4-layer stacks, because binary masks encode occupancy rather than layer count. Top RGB-D can also remain ambiguous under self-occlusion; an inward fold of a layered skirt may lie beneath a nearly unchanged visible surface. Bottom sensing reduces this ambiguity but requires a transparent support surface. The flat-folded comparison (Section 4.6) tests unfolding from a state whose support residual is already low; it shows neither that f_f or contact identify every folded state nor that bottom sensing alone resolves all layering ambiguity. Using bottom sensing only for privileged supervision, with top RGB-D plus history or active interaction at inference, remains future work.

The physical setup also limits this connection between perception and action. The dual-arm system has a restricted reachable workspace, handled by a left/right partition and an obstacle-avoidance heuristic, so some desired manipulation points cannot be reached exactly. Physical trials require a human operator to reset the garment and still involve many pick-and-pull interactions per trial, and RGB segmentation in both views is sensitive to illumination: the original setup used blackout curtains, and the lighting evaluation characterizes degradation outside those conditions. This study motivates organizing perception around physical relationships that distinguish useful actions. For cloth unfolding, support cues provide a link from visual observation to learned target selection.

AI USE STATEMENT

In this work, we used generative AI tools to implement methods, to design and provide feedback on experiments, to propose and refine hypotheses, to help develop the conceptual framework, and to interpret results. We have not used generative AI tools to generate synthetic data sets, to assist with translation, or to clean and reformat data sets; the remaining tasks requiring disclosure are not applicable to this work. Additionally, we used generative AI tools to create and modify figures, to create and edit code, to suggest experimental parameters, to brainstorm, and to draft, structure, and edit the paper. We have reviewed all AI-assisted work: AI-assisted code was reviewed by the authors before use, reported values were taken from the experiment logs, and all AI-assisted text and figures were checked and revised by the authors. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The appendix specifies the observation construction, controller parameters, checkpoint selection, scorer update, and per-seed results. The validation appendices state their evaluation protocols and report ranges, bounds, and uncertainty notation; Appendix V states the per-state physical evaluation protocol behind Table 5.

REFERENCES

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3):235–256, 2002.

Yahav Avigal, Lars Berscheid, Tamim Asfour, Torsten Kröger, and Ken Goldberg. Speedfolding: Learning efficient bimanual folding of garments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1–8, 2022. doi: 10.1109/IROS47612.2022.9981402.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025.

Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, Jie Lei, Tengyu Ma, Baishan Guo, Arpit Kalla, Markus Marks, Joseph Greer, Meng Wang, Peize Sun, Roman Rädle, Triantafyllos Afouras, Effrosyni Mavroudi, Katherine Xu, Tsung-Han Wu, Yu Zhou, Liliane Momeni, Rishi Hazra, Shuangrui Ding, Sagar Vaze, Francois Porcher, Feng Li, Siyuan Li, Aishwarya Kamath, Ho Kei Cheng, Piotr Dollár, Nikhila Ravi, Kate Saenko, Pengchuan Zhang, and Christoph Feichtenhofer. SAM 3: Segment Anything with Concepts, 2025. URL `https://arxiv.org/abs/2511.16719`.

Cheng Chi and Shuran Song. Garmentnets: Category-level pose estimation for garments via canonical space shape completion. In *IEEE/CVF International Conference on Computer Vision*, pp. 3324–3333, 2021.

Marco Cusumano-Towner, Arjun Singh, Stephen Miller, James F. O'Brien, and Pieter Abbeel. Bringing clothing into desired configurations with limited perception. In *IEEE International Conference on Robotics and Automation*, pp. 3893–3900, 2011. doi: 10.1109/ICRA.2011.5980327.

Peter Florence, Lucas Manuelli, and Russ Tedrake. Dense object nets: Learning dense visual object descriptors by and for robotic manipulation. In *Proceedings of the 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pp. 373–385, 2018.

Scott Fujimoto, Herke van Hoof, and David Meger. Addressing function approximation error in actor-critic methods. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1587–1596, 2018.

Irene Garcia-Camacho, Martina Lippi, Michael C. Welle, Hang Yin, Rika Antonova, Anastasiia Varava, Júlia Borràs, Carme Torras, Alessandro Marino, Guillem Alenyà, and Danica Kragic. Benchmarking bimanual cloth manipulation. *IEEE Robotics and Automation Letters*, 5(2):1111–1118, 2020. doi: 10.1109/LRA.2020.2965891.

Huy Ha and Shuran Song. Flingbot: The unreasonable effectiveness of dynamic manipulation for cloth unfolding. In *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pp. 24–33, 2022.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1861–1870, 2018.

Zixuan Huang, Xingyu Lin, and David Held. Mesh-based dynamics with occlusion reasoning for cloth manipulation. In *Proceedings of Robotics: Science and Systems*, New York City, NY, USA, June 2022. doi: 10.15607/RSS.2022.XVIII.011.

Biao Jia, Zherong Pan, Zhe Hu, Jia Pan, and Dinesh Manocha. Cloth manipulation using random-forest-based imitation learning. *IEEE Robotics and Automation Letters*, 4(2):2086–2093, 2019. doi: 10.1109/LRA.2019.2897370.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 661–670, 2010. doi: 10.1145/1772690.1772758.

Xingyu Lin, Yufei Wang, Jake Olkin, and David Held. Softgym: Benchmarking deep reinforcement learning for deformable object manipulation. In *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pp. 432–448, 2021.

Xingyu Lin, Yufei Wang, Zixuan Huang, and David Held. Learning visible connectivity dynamics for cloth smoothing. In *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pp. 256–266, 2022.

Alberta Longhini, Yufei Wang, Irene Garcia-Camacho, David Blanco-Mulero, Marco Moletta, Michael Welle, Guillem Alenyà, Hang Yin, Zackory Erickson, David Held, Júlia Borràs, and Danica Kragic. Unfolding the literature: A review of robotic cloth manipulation. *Annual Review of Control, Robotics, and Autonomous Systems*, 8:295–322, 2025. doi: 10.1146/annurev-control-022723-033252.

Jeremy Maitin-Shepard, Marco Cusumano-Towner, Jinna Lei, and Pieter Abbeel. Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In *IEEE International Conference on Robotics and Automation*, pp. 2308–2315, 2010.

Jan Matas, Stephen James, and Andrew J. Davison. Sim-to-real reinforcement learning for deformable object manipulation. In *Proceedings of the 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pp. 734–743, 2018.

Jose Sanchez, Juan-Antonio Corrales, Belhassen-Chedli Bouzgarrou, and Youcef Mezouar. Robotic manipulation and sensing of deformable objects in domestic and industrial applications: A survey. *The International Journal of Robotics Research*, 37(7):688–716, 2018. doi: 10.1177/0278364918779698.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Daniel Seita, Aditya Ganapathi, Ryan Hoque, Minho Hwang, Edward Cen, Ajay Kumar Tanwani, Ashwin Balakrishna, Brijen Thananjeyan, Jeffrey Ichnowski, Nawid Jamali, Katsu Yamane, Soshi Iba, John Canny, and Ken Goldberg. Deep imitation learning of sequential fabric smoothing from an algorithmic supervisor. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020.

Thomas Weng, Sujay Man Bajracharya, Yufei Wang, Khush Agrawal, and David Held. Fabricflownet: Bimanual cloth manipulation with a flow-based policy. In *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pp. 192–202, 2022.

Zhenjia Xu, Cheng Chi, Benjamin Burchfiel, Eric Cousineau, Siyuan Feng, and Shuran Song. Dextairity: Deformable manipulation can be a breeze. In *Proceedings of Robotics: Science and Systems*, New York City, NY, USA, June 2022. doi: 10.15607/RSS.2022.XVIII.017.

Andy Zeng, Shuran Song, Stefan Welker, Johnny Lee, Alberto Rodriguez, and Thomas Funkhouser. Learning synergies between pushing and grasping with self-supervised deep reinforcement learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4238–4245, 2018. doi: 10.1109/IROS.2018.8593986.

Andy Zeng, Pete Florence, Jonathan Tompson, Stefan Welker, Jonathan Chien, Maria Attarian, Travis Armstrong, Ivan Krasin, Dan Duong, Vikas Sindhwani, and Johnny Lee. Transporter networks: Rearranging the visual world for robotic manipulation. In *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pp. 726–747, 2021.

A Extended Experimental Diagnostics

Validation scope. Appendix studies examine the behavior, training horizon, and sensing assumptions behind the main comparisons. Qualitative rollouts from the same crumpled state make the principal failure mode visible: the top-only policy can increase projected area while leaving a folded front layer unresolved, whereas the reward-bandit affordance policy keeps both arms engaged and reduces the final folded residue from $F \approx 0.017$ to $F \approx 0.004$ with contact near 0.98. Mid-budget validation runs test whether the short scorer ablation is a pure horizon artifact: a 75k-step reward-bandit versus heuristic-prior comparison preserves the scorer advantage with a 0.168 within-table score gain, -0.0021 z-std change, and -0.0033 folded-area change. The offline mask-sensitivity probe further shows that threshold and morphology choices are relatively stable, while calibration-style shifts are the dominant failure: a one-pixel bottom-mask shift drops frontier IoU to 0.065 and moves picks by 23.2 cm. These diagnostics motivate the robustness tests below and are reported in full in Appendices D, L, and M.

A.1 Bottom-Mask Noise

The scorer ablation and the fold/contact–flatness associations are shown in Figures 5 and 19 of the main text; Table 6 reports the policy-level bottom-mask noise test.

Table 6: Policy-level bottom-mask noise robustness. Metrics use clean simulator geometry; only the policy's bottom-mask input is perturbed. "Score" is the within-panel composite S; absolute fold/contact/z-std columns should be used when comparing across rows. S is min–max normalized over the per-episode values of the rows in this table only and is not comparable to S in other tables. Both policies are trained for 25k steps with 3 seeds; clean-trained rows pool 4 evaluation episodes per seed and noise level, noise-trained rows pool 6.

Policy	Bottom-mask noise	Score	Return	Area	Fold ↓	Contact ↑	Z-std ↓
Reward-bandit affordance (clean-trained)	Clean	0.765 ± 0.035	32.8 ± 3.3	0.218 ± 0.005	0.0109 ± 0.0037	0.944 ± 0.018	0.0220 ± 0.0018
Reward-bandit affordance (clean-trained)	FN 5%	0.770 ± 0.016	30.9 ± 2.5	0.214 ± 0.004	0.0122 ± 0.0011	0.935 ± 0.005	0.0169 ± 0.0008
Reward-bandit affordance (clean-trained)	Flip 2%	0.780 ± 0.023	32.0 ± 3.6	0.215 ± 0.006	0.0106 ± 0.0022	0.945 ± 0.011	0.0183 ± 0.0011
Reward-bandit affordance (clean-trained)	1-px shift	0.563 ± 0.023	-409.7 ± 241.5	0.203 ± 0.005	0.0291 ± 0.0015	0.833 ± 0.008	0.0220 ± 0.0013
Reward-bandit affordance (clean-trained)	Mixed	0.554 ± 0.016	-283.1 ± 260.4	0.221 ± 0.004	0.0383 ± 0.0018	0.800 ± 0.008	0.0212 ± 0.0014
Reward-bandit affordance (noise-trained)	Clean	0.730 ± 0.029	22.7 ± 2.5	0.196 ± 0.004	0.0077 ± 0.0017	0.953 ± 0.011	0.0210 ± 0.0014
Reward-bandit affordance (noise-trained)	FN 5%	0.616 ± 0.040	21.2 ± 3.5	0.189 ± 0.006	0.0164 ± 0.0017	0.894 ± 0.014	0.0234 ± 0.0021
Reward-bandit affordance (noise-trained)	Flip 2%	0.670 ± 0.039	20.5 ± 3.3	0.192 ± 0.007	0.0119 ± 0.0017	0.924 ± 0.012	0.0229 ± 0.0017
Reward-bandit affordance (noise-trained)	1-px shift	0.451 ± 0.030	-2.4 ± 18.9	0.177 ± 0.009	0.0283 ± 0.0028	0.815 ± 0.014	0.0275 ± 0.0016
Reward-bandit affordance (noise-trained)	Mixed	0.410 ± 0.028	21.9 ± 3.6	0.183 ± 0.008	0.0363 ± 0.0019	0.768 ± 0.009	0.0255 ± 0.0016

B Extended Related Work

Robotic cloth manipulation is difficult because cloth has high-dimensional state, self-contact, and partial observability; surveys emphasize that methods often succeed by exploiting task structure in perception, modeling, or control (Sanchez et al., 2018; Longhini et al., 2025). Early systems used geometric cues and structured manipulation sequences for towel folding or clothing configuration estimation (Maitin-Shepard et al., 2010; Cusumano-Towner et al., 2011), while benchmarks for bimanual cloth manipulation formalized tasks such as spreading, folding, and dressing with task-specific quality measures (Garcia-Camacho et al., 2020). Learning-based approaches have since used imitation, scripted supervisors, and sim-to-real training to reduce manual policy design (Jia et al., 2019; Matas et al., 2018; Seita et al., 2020). This emphasis on task structure motivates our use of explicit support cues in the policy observation.

Recent cloth-unfolding and smoothing methods show the value of strong visual and action priors. Fabric smoothing has been learned from algorithmic supervision (Seita et al., 2020); dynamic dual-arm flinging can rapidly increase coverage (Ha & Song, 2022); flow-based bimanual policies use visual correspondence as both an observation and action representation (Weng et al., 2022); and efficient folding systems combine learned bimanual primitives with downstream folding routines (Avigal et al., 2022). Other work changes the actuation itself, for example using airflow to apply dense forces during deformable-object manipulation (Xu et al., 2022). Our work connects visual and action priors through contact-based dual-arm pulling: a transparent table supplies a bottom contact mask that informs both the state representation and manipulation targets.

Partial observability is a central obstacle for cloth state estimation. SoftGym provides deformable-object simulation tasks for learning-based manipulation (Lin et al., 2021), while visible-connectivity dynamics infer graph structure from partial point clouds for cloth smoothing (Lin et al., 2022). GarmentNets addresses category-level garment pose estimation through canonical-space shape completion (Chi & Song, 2021), and MEDOR explicitly reasons about occluded cloth regions by reconstructing a mesh and refining it with self-supervised test-time losses (Huang et al., 2022). Our representation focuses on the support relation needed for action without recovering a full 3D cloth state: the bottom mask indicates which top-visible cells are grounded and where the estimated folded region meets contact.

Deep RL algorithms such as SAC, TD3, and PPO provide general continuous-control baselines (Haarnoja et al., 2018; Fujimoto et al., 2018; Schulman et al., 2017), and affordance-oriented manipulation shows that action-relevant spatial structure can improve exploration and control (Zeng et al., 2018; Florence et al., 2018). Transporter networks further show that spatial action parameterizations and visual correspondences can make manipulation policies more sample efficient (Zeng et al., 2021). For cloth unfolding, the main difficulty is not only action dimensionality but also misleading top-view progress: projected area can increase while hidden folds remain stacked. We therefore use bottom-view fold-contact frontiers both as observation channels and as residual action-prior targets, leaving SAC to learn continuous corrections rather than replacing the policy with a scripted controller.

The target-ranking component draws on contextual bandits and lightweight online adaptation, where action preferences are updated from observed feedback and UCB-style bonuses encourage uncertain choices (Li et al., 2010; Auer et al., 2002). We use these ideas narrowly: a linear scorer ranks local frontier cells from fold/contact/edge features and is updated from bounded reward or self-supervised progress. It does not model long-horizon cloth dynamics or claim a regret guarantee; its purpose is to adapt the target-selection part of an interpretable bottom-view action prior.

C HARDWARE SETUP

The transparent-table prototype uses two Intel RealSense D435 cameras: an overhead camera observing the cloth from above and an upward-facing camera mounted below the workspace to observe contact through the transparent support surface. The support surface is an off-the-shelf tempered-glass table that can be purchased commercially, rather than a custom optical bench. Dual-arm manipulation is performed with two Agile Nero robot arms positioned on opposite sides of the table. Figure 8 shows the physical layout.

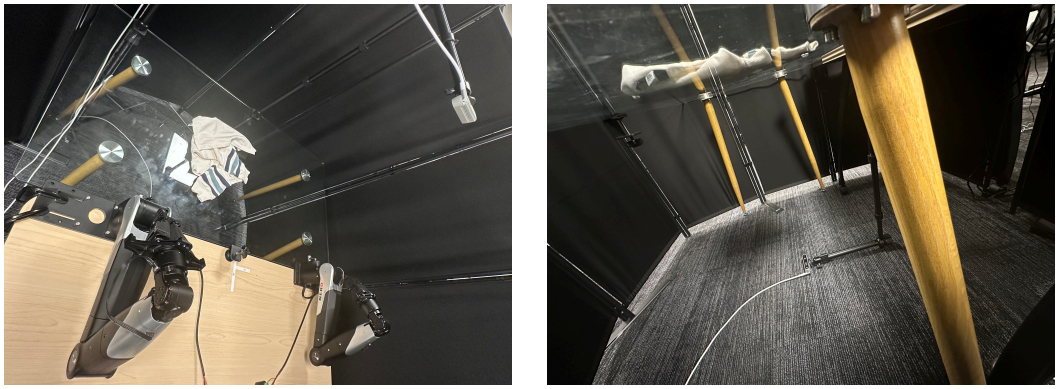


Figure 8: Physical transparent-table setup. Two Intel RealSense D435 cameras provide top and bottom observations through a commercially available tempered-glass table, while two Agile Nero arms perform dual-arm cloth manipulation.

D QUALITATIVE METHOD COMPARISON SNAPSHOTS

The two policies in Section 4.4 are easiest to interpret side-by-side. Figure 6 shows deterministic rollouts of the top-only SAC baseline (top row) and the reward-bandit affordance policy (bottom row) starting from the identical crumpled cloth state used in the main simulation studies. The two policies share the simulator seed and the same cloth-init snapshot, so any visible difference across the row is attributable to the policy and observation choice rather than to initial conditions. Cloth particles are coloured by height above the table; the red and blue spheres mark the left and right end-effectors.

Two qualitative differences are visible. First, the top-only policy never recovers the folded layer that sits along the front of the garment, while the reward-bandit affordance policy progressively flattens it as the episode advances. Second, the top-only policy stalls with the right gripper held outside the cloth (a common failure mode that the residual frontier-target prior is designed to avoid). The final-frame folded mask area F and contact ratio C quantify the gap: the reward-bandit affordance policy reaches $F \approx 0.004$ and $C \approx 0.98$ at $t{=}T$, versus $F \approx 0.017$ and $C \approx 0.91$ for the top-only baseline, which is consistent with the aggregated numbers in Table 3.

E REAL-WORLD EXPERIMENTAL RECORDS

Figures 9 and 10 show eight real-world experimental records collected with the transparent-table setup across four garments. Figure 7 visualizes one representative trial with aligned progress metrics.

During execution, the system records synchronized RGB observations, robot states, segmentation outputs, and reward-related metrics. The top row of Figure 7 shows key moments from the execution sequence. The garment is segmented using SAM 3 (Carion et al., 2025), and the segmentation masks provide a direct visual indication of the observed garment state. Across the sequence, the robot progressively changes the garment configuration through physical interaction, producing a more favorable and observable cloth arrangement.

The bottom plot reports normalized task-relevant metrics extracted from the same record. To make the trend interpretable under real-world sensor noise, the plot keeps small local fluctuations while

showing an uncertainty band. The plotted points are aligned with the keyframes above, so each image corresponds to the metric value at the same stage of execution. The overall upward trend suggests that the robot action sequence improves the measured garment state over time. This result indicates that the visual perception pipeline provides stable enough feedback for evaluating task progress despite occlusions, lighting variation, and cloth deformation.

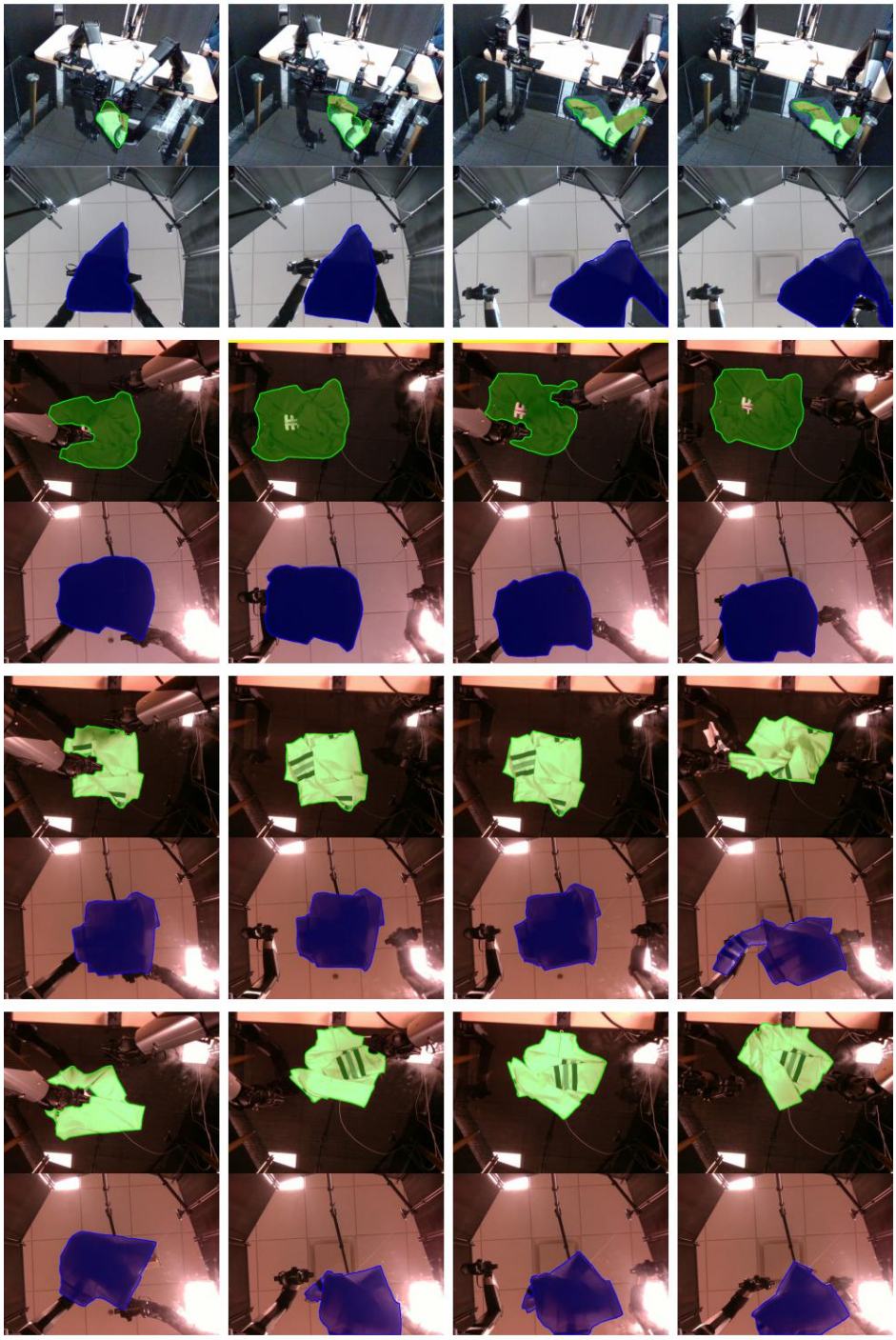


Figure 9: Real-world experimental records from the transparent-table setup (upper half). The full record contains eight trials across four garments. In each trial block, the top image is the top view and the bottom image is the bottom view.

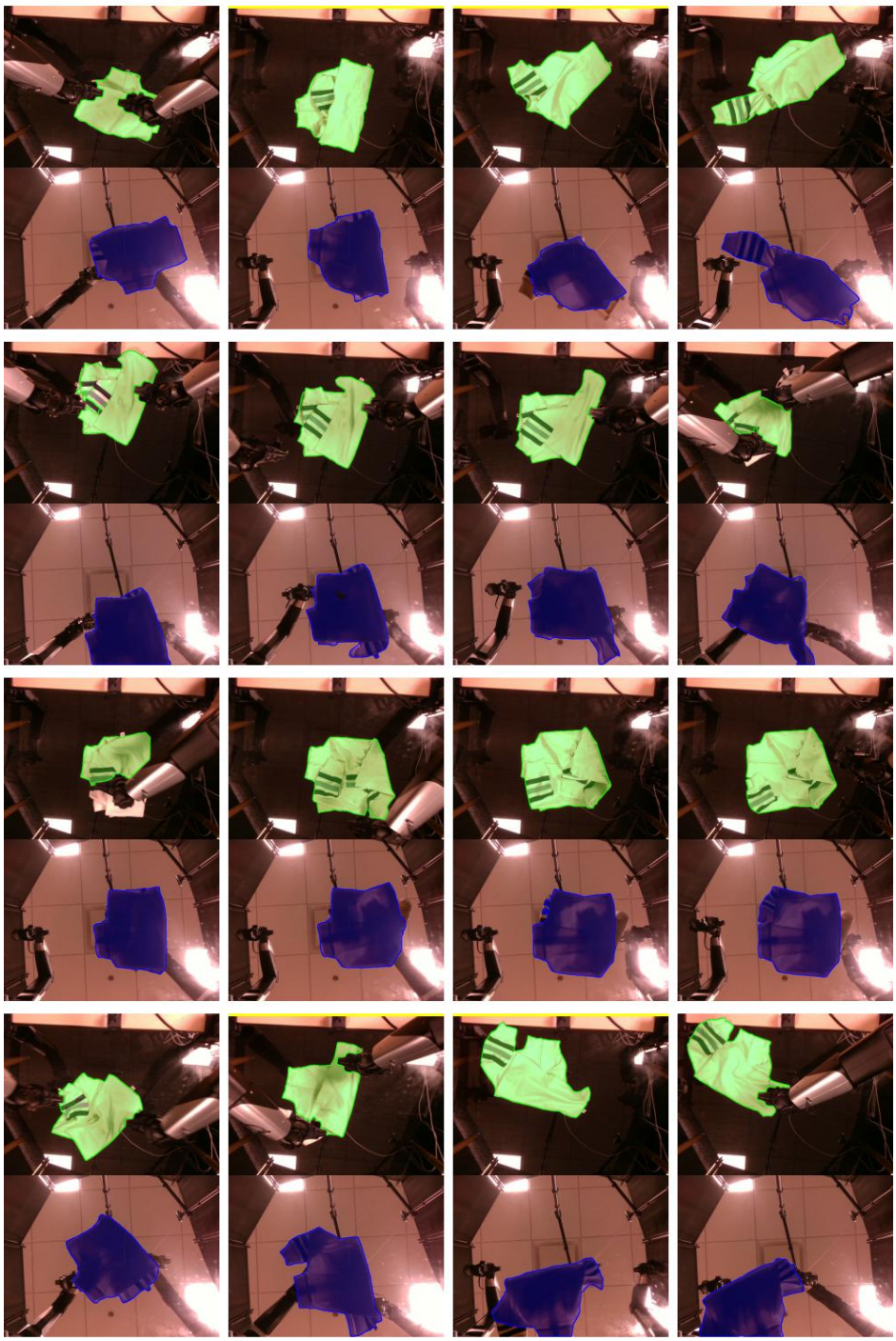


Figure 10: Real-world experimental records from the transparent-table setup (lower half), continued from Figure 9.

F Implementation Details

The SAC implementation uses twin critics, automatic entropy tuning, replay updates after a random-action warmup, and deterministic evaluation rollouts for checkpoint selection. The auto-unfold preset selects checkpoints by deterministic evaluation: the long top/raw-dual runs use `eval_success_rate` with six evaluation episodes every 2000 updates, while the short scorer and occlusion studies use `eval_avg_return` with four evaluation episodes every 1000 updates. The main simulation tables are produced from `latest_metrics.json` or from explicit post-hoc evaluation scripts, and incomplete runs are excluded by the finalizers.

The mask grid is 32×32 over a square table-centered ROI with half extent 0.40 m. Simulated SAM-style noise uses false-negative probability 0.01 and random flip probability 0.001. The bottom mask threshold is $\epsilon_{bot} = 0.035$ m above the table. The bottom mask is computed from cloth particles and table height, approximating an upward-looking transparent-table camera; it does not model calibration error, refraction/reflection, camera exposure, or learned segmentation bias.

Morphology is fixed unless stated otherwise. Both fold/contact frontier dilation and silhouette-edge erosion use one iteration of a 3×3 square binary kernel. The frontier is $\mathrm{dilate}(M^{fold})\,\mathrm{dilate}(M^{contact})M^{top}$. The residual prior uses pull distance 0.06 m in the reward-bandit runs, pick tolerance 0.045 m, prior weight $\lambda = 0.35$, and side-split candidate sets relative to the garment center. The initial linear scorer weights over [frontier, edge, fold, top, contact, outward] are $[4.0, 1.5, 1.0, 0.4, -0.2, 0.35]$; learned weights are clipped to $[0, 8] \times [0, 4] \times [0, 4] \times [0, 2] \times [-2, 2] \times [-1, 2]$. Reward-bandit runs use learning rate 0.20, exploration $\epsilon = 0.10$ over top-quartile candidate cells, baseline EMA coefficient 0.90, and weight decay 0.01 toward the heuristic initialization.

The visual encoders use feature dimension 128 and channel widths $(32, 64, 96)$. The top-mask encoder applies three convolutional blocks with stride 2 followed by global average pooling. The raw dual encoder builds separate top and bottom stems, tokenizes the final 8×8 feature maps into 64 tokens each, applies four-head top-to-bottom and bottom-to-top cross attention, and pools the original and attended tokens. The affordance encoder uses a six-channel global stem, a two-channel fold/frontier stem, and a two-channel contact/edge stem, then applies the same bidirectional four-head attention and pooling before concatenating the visual feature with proprioception.

G Perception Validation

Figure 11 qualitatively validates the perception representation used by the real-world policy. The bottom mask, from bottom-view segmentation plus a depth-based near-table height test on the visible cloth underside, captures cloth resting on or near the glass, while the top mask, from RGB segmentation alone, captures the full visible garment silhouette. Combining these two views separates likely folded-over cloth (M_{fold}) from table-contacting cloth (M_{contact}). The resulting frontier mask consistently lies near visually salient fold boundaries, which are the regions where pulling is expected to expose hidden cloth and increase garment coverage. The selected pick points are sampled from high-scoring frontier regions and generally fall on or near these estimated fold-contact boundaries, supporting the use of bottom-camera contact perception as a practical signal for real-world cloth unfolding.

G.1 SAM3 Garment Segmentation Validation

To validate the robustness of the perception module across different garment appearances, we collected 12 real-world examples using the dual RealSense camera setup. For each example, the two RGB streams were vertically stitched and passed to SAM3 with the text prompt "cloth". We then applied a per-view post-processing step that keeps the highest-confidence denoised garment mask in each camera view. The resulting visualization is shown in Figure 12, where each row contains the raw stitched RGB observation, the SAM3 overlay, and the final binary mask.

Across the collected samples, SAM3 successfully detected the garment in both camera views for all 12 cases. The examples cover visually distinct clothing items, including dark, light, colored, patterned, and highly deformable garments. The detection confidence over the 24 retained per-view masks ranged from approximately 0.65 to 0.96, with an average score of about 0.85. These results

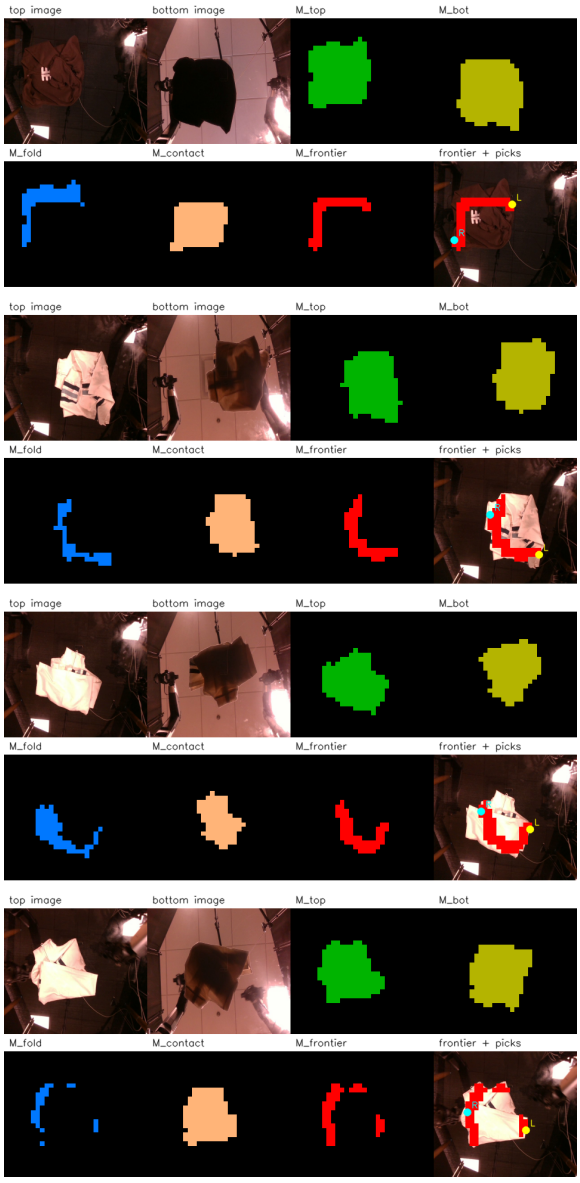


Figure 11: Perception validation for bottom-contact mask and fold-contact frontier in real-world scenes. Each row shows one recorded cloth state from the real robot setup. From left to right, the panels show the top RGB image, bottom RGB image through the transparent table, top cloth mask M_{top}, bottom contact mask M_{bot}, residual folded-cloth mask $M_{\text{fold}} = M_{\text{top}} - M_{\text{bot}}$, contact mask $M_{\text{contact}} = M_{\text{top}} \cap M_{\text{bot}}$, fold-contact frontier M_{frontier}, and the selected left/right pick points overlaid on the top image. The red frontier indicates the estimated boundary between upper folded cloth and lower table-contacting cloth. Yellow and cyan dots denote the selected grasp candidates for the two arms.

indicate that the text-prompted SAM3 perception pipeline provides sufficiently stable garment masks for downstream real-world manipulation and paper-quality qualitative visualization.

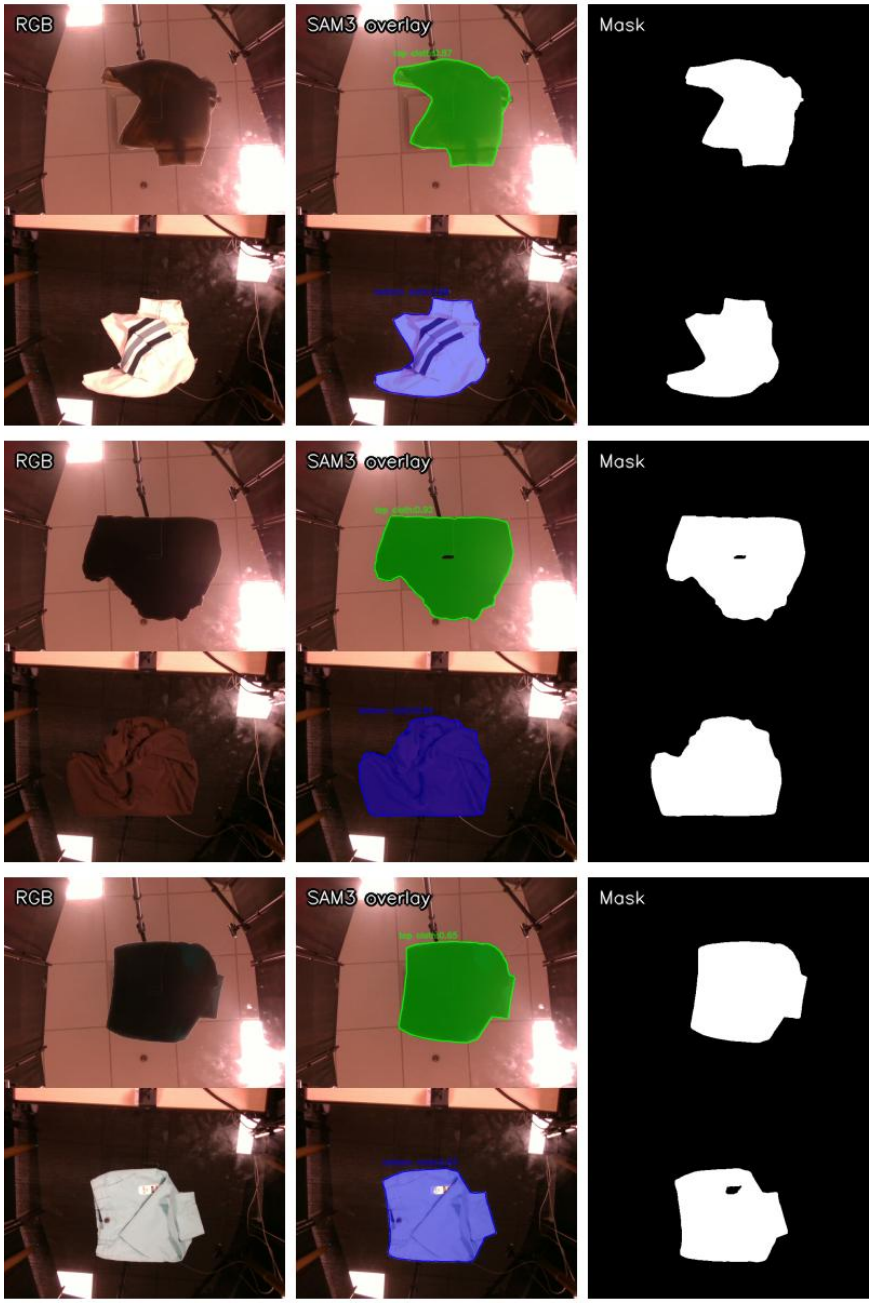


Figure 12: **Qualitative validation of SAM3 garment segmentation across real-world clothing instances.** Each row shows one captured garment configuration from the dual RealSense setup. The three columns show the stitched RGB observation, the SAM3 segmentation overlay, and the extracted binary garment mask, respectively. Using the text prompt "cloth", SAM3 consistently localizes the target garment in both camera views across 12 clothing appearances with different colors, textures, shapes, and degrees of deformation. This test evaluates bottom-view recognition of garments placed on a glass table. The per-view detections were filtered to retain one garment mask per camera view.

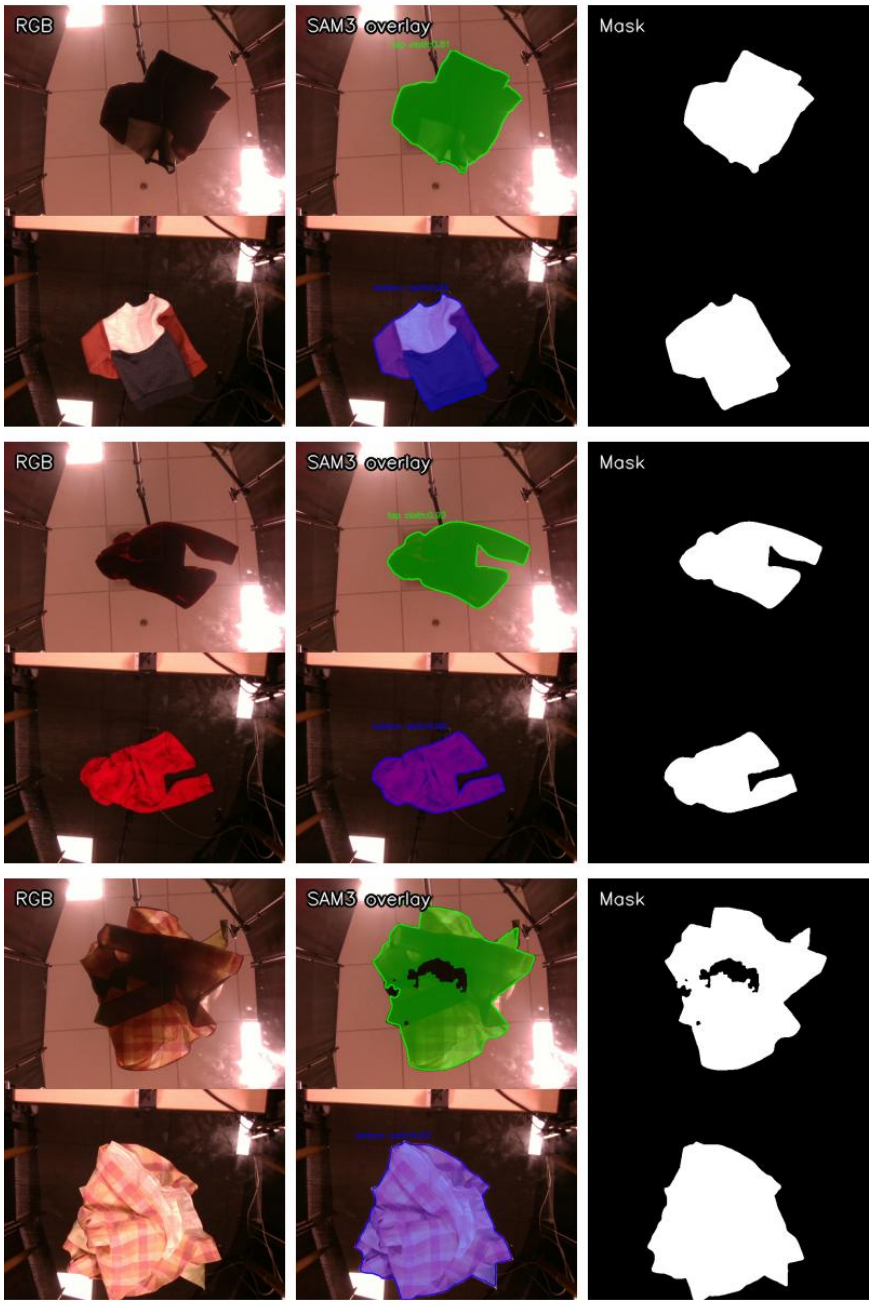


Figure 13: Qualitative validation of SAM3 garment segmentation across real-world clothing instances, continued from Figure 12.

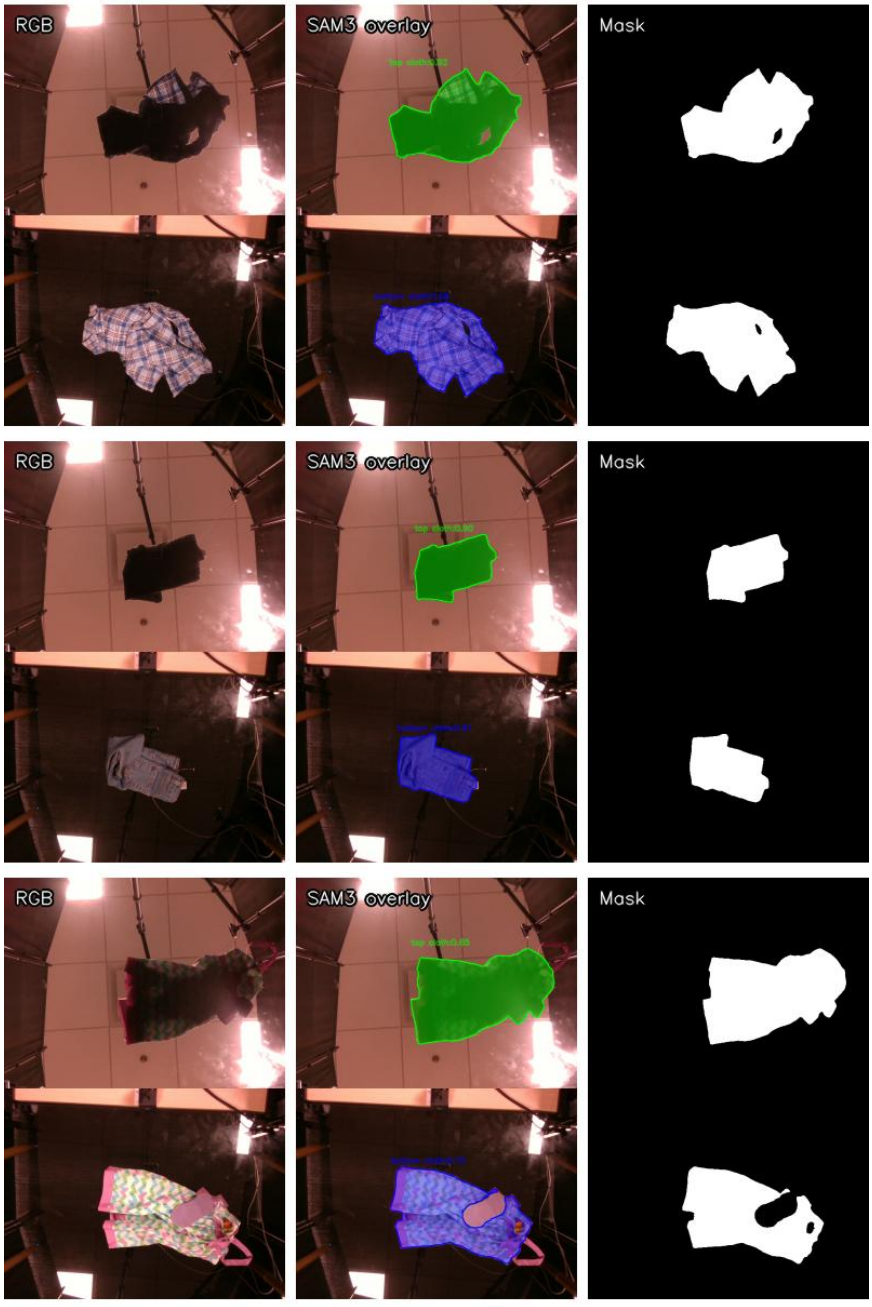


Figure 14: Qualitative validation of SAM3 garment segmentation across real-world clothing instances, continued from Figure 12.

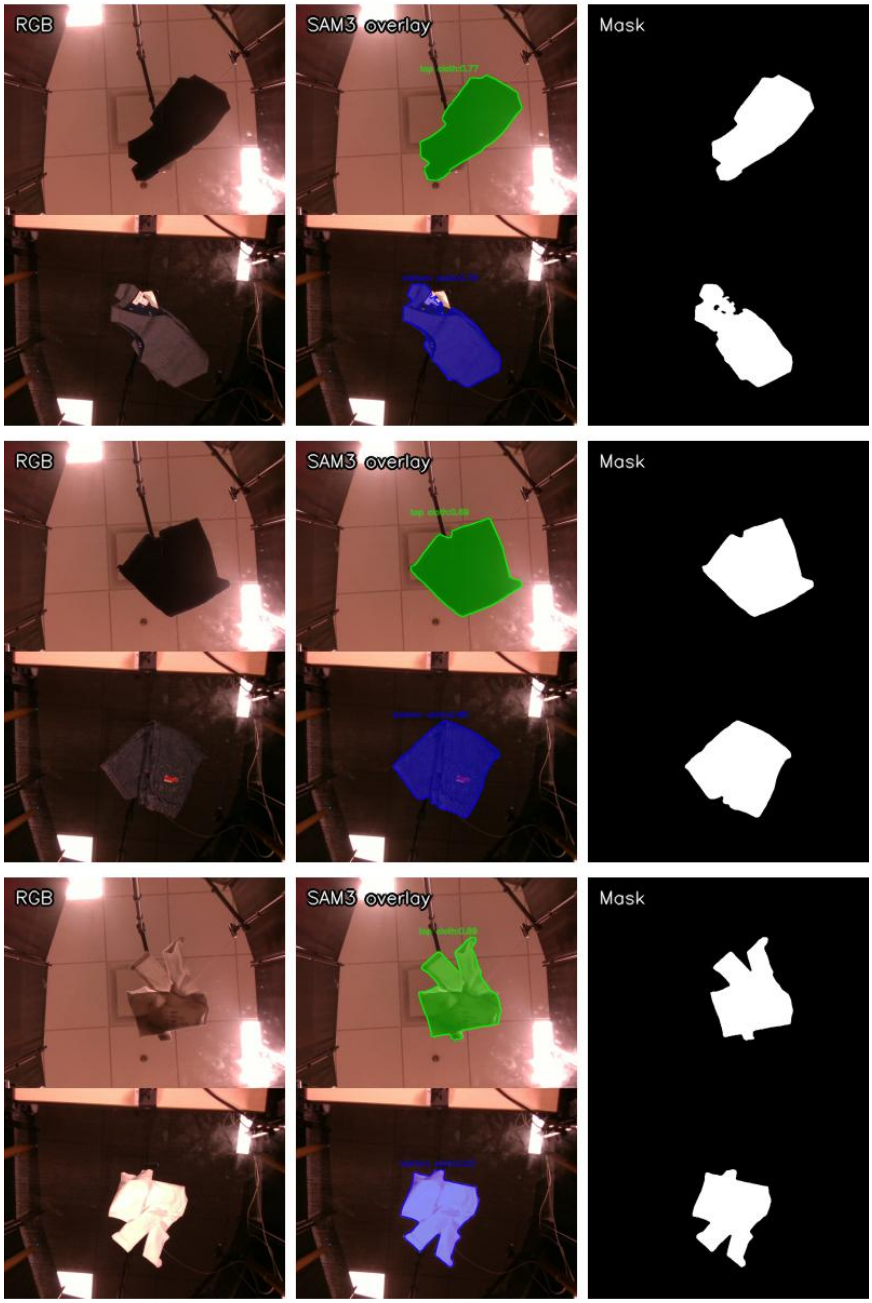


Figure 15: Qualitative validation of SAM3 garment segmentation across real-world clothing instances, continued from Figure 12.

H LONG-HORIZON TOP VS. RAW DUAL-VIEW RUNS

Table 7: One-million-step SAC runs used to evaluate whether raw bottom-view pixels add information beyond a top-only silhouette.

Condition	Obs.	Fusion	Episodes	Success	Return	Final area	Best area	Spread	Z-std
Top-mask	top_mask	top_cnn	2493	0.00	13.7 ± 2.0	0.200 ± 0.000	0.216 ± 0.000	0.183 ± 0.000	0.0271 ± 0.0003
Dual-view	dual_mask	cross_attention	2493	0.00	14.7 ± 1.8	0.202 ± 0.000	0.217 ± 0.000	0.184 ± 0.000	0.0218 ± 0.0001

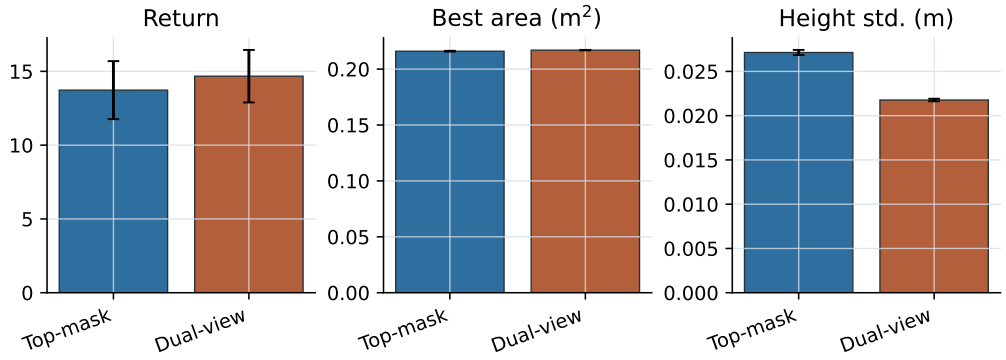


Figure 16: Aggregate episode statistics for the long top-mask and raw dual-view runs.

I ACTION-PRIOR PROBES

Table 8: No-learning 80-step bottom-view action-prior probe. Positive Δ area, Δ fold, Δ contact, and Δ z-std mean area increase, fold reduction, contact-ratio increase, and height-std reduction, respectively.

Condition	Score	Δ area	Δ fold	Δ contact	Δ z-std
Affordance prior	0.5414	0.0469	-0.0025	-0.0094	0.0059
Random dual	0.5237	0.0052	0.0025	0.0163	0.0071
Zero dual	0.5030	0.0072	0.0019	0.0120	0.0004
Affordance prior + jitter	0.4974	0.0204	-0.0069	-0.0312	0.0051

The prior is heuristic by construction. It depends on the binary morphology above, the side split, the pick tolerance, and the outward pull distance. Existing probes show that the prior can increase projected area without learning, but also that aggressive pulling can reduce contact or increase folding. We therefore treat the residual prior as a proposal generator for SAC, not as a complete controller.

Table 9: Lightweight sensitivity probe over outward pull distance, using two 80-step no-learning rollouts per setting. Negative Δ fold or Δ contact indicates that the prior increased folded area or reduced contact, respectively. Short pulls are safer than the default long pull.

Pull setting	Pull (m)	Score	Δ area	Δ fold	Δ contact	Final z-std
Short	0.06	0.521	0.0388	-0.0056	-0.0236	0.0181
Medium	0.09	0.509	0.0171	-0.0016	-0.0074	0.0182
Default	0.12	0.445	0.0063	-0.0138	-0.0724	0.0244

Table 10: Twelve-hour residual-prior comparison. These runs stopped before the requested 280k-step budget but completed about 248k steps and are used only as an exploratory diagnostic.

Variant	Steps	Score	Eval return	Area	Fold ↓	Contact ↑
Affordance residual prior	248598	0.650	18.7	0.199	0.0113	0.936
Affordance observation	248526	0.350	33.9	0.214	0.0141	0.924

J Discussion

The experiments distinguish access to support information from its use in control. A bottom contact mask complements the overhead silhouette; the fold-contact representation makes their relationship explicit; and the learned scorer adapts target selection from interaction. This sequence explains why the evaluation tracks several aspects of unfolding: final area, best area, spread, folded area, contact ratio, and height standard deviation. Together, these measures distinguish visible coverage from residual folds and grounded support. Binary success remains meaningful under a fixed threshold, but its task-dependent tolerance makes it secondary to these continuous diagnostics. The main simulation tables are drawn from run directories, with incomplete or failed experiments excluded by the finalization script.

K Robustness Diagnostics

Table 11 and Figure 17 give the full top-camera occlusion stress test behind the main-text summary. Bottom-view completion fills top-camera holes using the bottom contact mask. Under arm-like occlusion, the reward-bandit affordance policy keeps the highest score and return and the lowest folded area; the gain is not in z-std. This stress test supports robustness to top-camera occlusion, while not modeling bottom-camera calibration or optical failures.

Table 11: Structured top-camera occlusion stress test. "Arm-like" uses two elongated end-effector-centered occluders; bottom completion is enabled only for the reward-bandit affordance row.

Policy	Eval view	Score	Return	Area	Fold ↓	Z-std ↓
Raw dual-view SAC	Clean	0.692 ± 0.070	14.2 ± 4.9	0.155 ± 0.025	0.0021 ± 0.0008	0.0182 ± 0.0014
Raw dual-view SAC	Arm-like top occl.	0.708 ± 0.079	19.9 ± 2.1	0.181 ± 0.015	0.0073 ± 0.0036	0.0211 ± 0.0034
Reward-bandit no completion	Clean	0.819 ± 0.032	38.0 ± 3.4	0.222 ± 0.006	0.0094 ± 0.0034	0.0194 ± 0.0025
Reward-bandit no completion	Arm-like top occl.	0.827 ± 0.022	35.4 ± 3.8	0.213 ± 0.005	0.0038 ± 0.0010	0.0214 ± 0.0023
Reward-bandit affordance	Clean	0.809 ± 0.025	22.2 ± 10.7	0.210 ± 0.007	0.0072 ± 0.0036	0.0185 ± 0.0015
Reward-bandit affordance	Arm-like top occl. + bottom completion	0.860 ± 0.023	36.5 ± 2.9	0.220 ± 0.005	0.0033 ± 0.0008	0.0196 ± 0.0024
Raw dual occlusion-trained	Clean	0.701 ± 0.050	17.5 ± 4.6	0.186 ± 0.009	0.0075 ± 0.0017	0.0261 ± 0.0030
Reward-bandit occlusion-trained	Clean	0.802 ± 0.044	32.5 ± 2.2	0.216 ± 0.004	0.0092 ± 0.0045	0.0201 ± 0.0036
Top-only occlusion-trained	Clean	0.540 ± 0.049	23.3 ± 4.0	0.220 ± 0.005	0.0347 ± 0.0051	0.0406 ± 0.0044
Raw dual occlusion-trained	Arm-like top occl.	0.723 ± 0.046	18.3 ± 4.9	0.187 ± 0.010	0.0047 ± 0.0007	0.0259 ± 0.0033
Reward-bandit occlusion-trained	Arm-like top occl. + bottom completion	0.829 ± 0.022	32.8 ± 3.7	0.223 ± 0.005	0.0072 ± 0.0017	0.0220 ± 0.0020
Top-only occlusion-trained	Arm-like top occl.	0.563 ± 0.056	23.0 ± 3.6	0.221 ± 0.006	0.0324 ± 0.0058	0.0404 ± 0.0045

L Extended Validation Runs

The main learning and scorer studies use completed one-million-step and 25k-step experiments. To address seed and horizon limitations, we also run a supplementary validation queue: three additional clean top-only/raw-dual seeds at 50k steps, and three reward-bandit versus three heuristic residual-prior seeds at 75k steps. Scores in Tables 12 and 13 are min-max normalized within each table only. At 50k steps over 3 top-only and 3 raw-dual seeds, raw dual changes z-std by -0.0263 relative to top-only and changes folded area by -0.0056. These runs are a mid-budget training-seed check rather than a replacement for a full multi-seed one-million-step study. At 75k steps over 3 reward-bandit seeds and 3 heuristic seeds, reward-bandit changes the within-table score by 0.168, z-std by -0.0021, and folded area by -0.0033 relative to the fixed scorer.

M Bottom-Mask Sensitivity Probe

To test sensitivity to ϵ_{bot}, morphology, side splitting, and bottom-mask failures without launching another full RL sweep, we collect 20 shared simulator states from three short crumpled-cloth rollouts

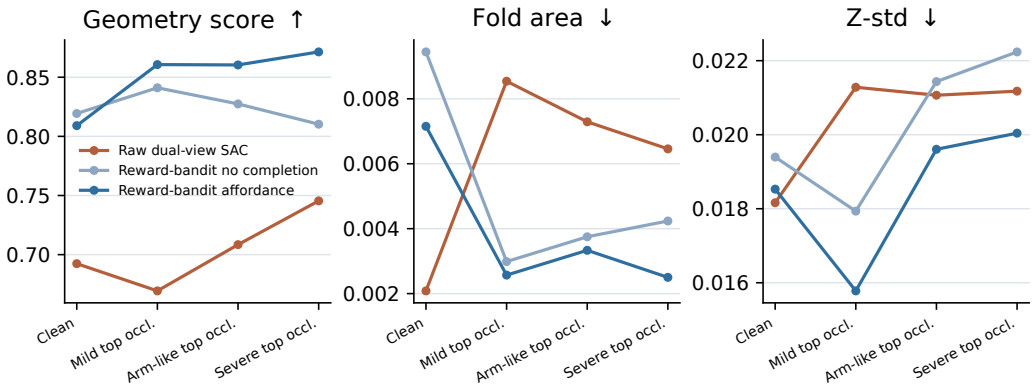


Figure 17: Occlusion stress test across increasing top-camera occluder sizes. The bottom-view reward-bandit affordance remains the strongest row under arm-like occlusion because the bottom contact mask supplies information not visible in the occluded top silhouette.

Table 12: Extended mid-budget top-only versus raw dual-view training-seed check. This table addresses whether the one-million-step pair is representative at shorter horizons; it does not replace full long-horizon multi-seed training.

Policy	Runs	Steps	Score	Return	Area	Fold ↓	Z-std ↓
Top-only SAC	3	50k	0.445 ± 0.142	23.9 ± 6.2	0.194 ± 0.031	0.0412 ± 0.0051	0.0681 ± 0.0295
Raw dual-view SAC	3	50k	0.579 ± 0.173	27.0 ± 8.1	0.204 ± 0.005	0.0356 ± 0.0170	0.0418 ± 0.0143

Table 13: Extended reward-bandit versus heuristic residual-prior check. The longer 75k-step horizon tests whether the 25k scorer conclusion remains visible after additional RL updates.

Policy	Runs	Steps	Score	Return	Area	Fold ↓	Z-std ↓
Heuristic residual prior	3	75k	0.506 ± 0.128	24.6 ± 2.6	0.203 ± 0.005	0.0121 ± 0.0045	0.0246 ± 0.0036
Reward-bandit residual prior	3	75k	0.675 ± 0.171	25.0 ± 3.5	0.206 ± 0.003	0.0088 ± 0.0031	0.0224 ± 0.0026

and re-compute the affordance masks and heuristic pick targets offline. The baseline is $\epsilon_{bot} = 3.5$ cm with a one-iteration 3×3 square kernel and the table-centered side split used by the code. The table reports bottom-mask IoU and frontier IoU relative to this baseline, mean pick displacement for the two arms, the fraction of missing picks, and fold/contact area changes normalized by top-mask area. We extend the scenarios to include $\epsilon_{bot} \in \{1.5, 2.0, 2.5, 4.5, 5.0\}$ cm, three percentile-based adaptive thresholds (p25, p35, p50), non-square dilation kernels ($1\times3, 3\times1, 1\times5, 5\times1, 5\times5, 7\times7$), a 3×3 kernel with two iterations, two side-split alternatives, and a broader noise family (5%/10% false negatives, 2% random flips, 1/2-pixel calibration shifts, block dropout, and a mixed failure).

The three takeaways are: *(i)* the offline frontier is largely insensitive to morphology and the precise ϵ_{bot} at and above 2.5 cm (frontier IoU ≥ 0.5, pick shift ≤ 5.4 cm); *(ii)* the choice of adaptive percentile in $\{25, 35, 50\}$ produces identical bottom and frontier masks on the shared states because the selected ϵ_{bot} is clipped to the same admissible interval; *(iii)* the dominant failure mode is calibration-style shift, which collapses frontier IoU to ≤ 0.07 and produces > 20 cm pick displacement even at one pixel. Policy-level evaluation of these perturbations is in Appendix N.

Table 14: Post-hoc sensitivity of bottom-view affordances and pick targets. This is a mask/proposal stability probe on shared states, not a policy retraining experiment. Low frontier IoU and large pick shift under shifted or mixed bottom-mask failures indicate that real calibration and segmentation quality are important assumptions.

Perturbation	Bottom IoU	Frontier IoU	Pick shift (cm)	Empty picks	Fold/contact Δ
$\epsilon_{bot} = 1.5$ cm	0.925	0.150	11.8	0.0%	0.074/-0.074
$\epsilon_{bot} = 2.0$ cm	0.967	0.307	9.3	0.0%	0.033/-0.033
$\epsilon_{bot} = 2.5$ cm	0.984	0.512	5.4	0.0%	0.015/-0.015
$\epsilon_{bot} = 4.5$ cm	0.993	0.668	2.7	0.0%	-0.007/0.007
$\epsilon_{bot} = 5.0$ cm	0.992	0.558	4.5	0.0%	-0.008/0.008
Adaptive p25	0.967	0.307	9.3	0.0%	0.033/-0.033
Adaptive p35	0.967	0.307	9.3	0.0%	0.033/-0.033
Adaptive p50	0.967	0.307	9.3	0.0%	0.033/-0.033
Kernel 1×3	1.000	0.536	0.5	0.0%	0.000/0.000
Kernel 3×1	1.000	0.403	0.8	0.0%	0.000/0.000
Kernel 1×5	1.000	0.489	0.2	0.0%	0.000/0.000
Kernel 5×1	1.000	0.401	0.6	0.0%	0.000/0.000
Kernel 5×5	1.000	0.508	0.2	0.0%	0.000/0.000
Kernel 7×7	1.000	0.344	0.4	0.0%	0.000/0.000
Kernel 3×3, 2 iters	1.000	0.508	0.2	0.0%	0.000/0.000
Top-centroid side split	1.000	1.000	0.0	0.0%	0.000/0.000
World-y side split	1.000	1.000	0.0	0.0%	0.000/0.000
Bottom 5% false negatives	0.949	0.123	10.2	0.0%	0.050/-0.050
Bottom 10% false negatives	0.897	0.078	12.5	0.0%	0.102/-0.102
Bottom 2% random flips	0.929	0.262	7.4	0.0%	0.020/-0.020
Bottom 1-pixel shift	0.758	0.065	23.2	0.0%	0.131/-0.131
Bottom 2-pixel shift	0.592	0.023	21.5	0.0%	0.248/-0.248
Bottom block dropout	0.951	0.566	4.1	0.0%	0.048/-0.048
Mixed bottom failure	0.752	0.061	11.2	0.0%	0.160/-0.160

N Bottom-Mask Noise Policy Evaluation

Table 6 reports the policy-level bottom-mask noise stress test in the main paper. For each clean-trained checkpoint we run four deterministic eval episodes per noise level, repeated across the three reward-bandit seeds. The noise levels match the offline probe semantics so the appendix can be cross-referenced with Table 14. The noise-trained rows are produced by retraining the reward-bandit affordance policy with the same configurable perturbation family applied during training (3 seeds, 25k steps, "mixed" noise on every frame: FN 5%, flip 1%, shift 1 px), then evaluated with the same scripted noise grid using six episodes per seed and noise level.

Two observations support the main-text robustness summary. First, the dominant policy-level degradation tracks the offline ranking from Appendix M: 1-pixel shift and mixed failure produce the largest drops in geometry score, FN 5% and flip 2% are mild, and contact drops substantially under shift and mixed failure ($0.944 \rightarrow 0.833$ and 0.800) and is essentially unchanged under FN 5% and flip 2%. Second, noise-augmented training removes the return collapse under shift and mixed noise

($-409.7 \rightarrow -2.4$ and $-283.1 \rightarrow 21.9$) but does not recover support geometry: contact is lower than the clean-trained policy under every perturbation (e.g., $0.800 \rightarrow 0.768$ under mixed), folded area under shift and mixed noise remains 0.028–0.036, and on clean input the retrained policy has lower area and return ($0.218 \rightarrow 0.196$; $32.8 \rightarrow 22.7$) with slightly lower fold and higher contact. We therefore read augmentation as preventing penalty accumulation from corrupted cues, not as a defense that restores cue quality; registration error is addressed physically in Appendix S.

O UCB-STYLE SCORER ABLATION

To address whether explicit uncertainty modeling improves the bandit-style update, we add a reward-UCB scorer variant that augments $s_t(i,j) = w_t^\top \phi_t(i,j)$ with a count-based exploration bonus $c\sqrt{\log(N{+}1)/(n_{ij}{+}1)}$ (Auer et al., 2002), where n_{ij} is the per-cell visit count and N the total picks since the last episode reset, and otherwise reuses the reward-bandit update rule, learning rate $\eta{=}0.20$, $\epsilon{=}0.10$ top-quartile exploration, and weight clipping. We run three matched seeds at 25k steps and report the result alongside reward-bandit in Table 15. The point of this ablation is not to introduce a new state-of-the-art scorer but to test whether replacing ϵ-greedy-on-top-quartile with an explicit visit-count bonus produces a measurable change at this scale; the rest of the pipeline is held fixed.

The two variants land in the same operating regime: reward-UCB has a slightly higher within-panel composite (0.607 vs 0.560), lower folded-mask area (0.0039 vs 0.0060), and modestly higher contact ratio (0.976 vs 0.968), but is essentially tied on coverage (0.203 vs 0.212 m^2) and z-std (0.0197 vs 0.0191 m), all within one standard error. This is consistent with the variance-control analysis in Appendix P.1: explicit visit-count uncertainty does not yield a measurable score improvement at the 25k-step scale used here, suggesting that the dominant variance source for this scorer family is feedback noise rather than under-exploration.

Table 15: Reward-bandit vs. reward-UCB scorer ablation at matched 3 seeds and 25k SAC steps. The UCB variant adds a $c\sqrt{\log(N{+}1)/(n_{ij}{+}1)}$ exploration bonus to the same linear score; all other hyperparameters are identical. "Score" is the within-panel composite; absolute fold/contact/ z-std columns are the most informative comparison.

Variant	Runs	Score	Area	Fold ↓	Contact ↑	Z-std ↓
Reward-bandit scorer	3	0.560 ± 0.230	0.212 ± 0.007	0.0060 ± 0.0018	0.968 ± 0.011	0.0191 ± 0.0031
Reward-UCB scorer	3	0.607 ± 0.188	0.203 ± 0.010	0.0039 ± 0.0008	0.976 ± 0.007	0.0197 ± 0.0015

P SCORER ABLATION DETAILS

P.1 SCORER UPDATE AND GATING PSEUDO-CODE

Algorithm 1 gives the high-level policy-side logic summarized in Section 3. It intentionally omits implementation details such as empty-set checks and UCB counters; the following paragraph lists the variance controls used by the update.

Algorithm 1 Reward-Bandit Scorer Update

initialize scorer weights at heuristic prior w_0 and EMA feedback baseline
for each environment step **do**
  build feasible per-arm target candidates from top-cloth cells
  rank candidates with the current affordance scorer
  fall back to the fixed heuristic if learned scores are uninformative
  select one target per arm with ϵ-greedy exploration
  execute the clipped residual prior action
  convert rollout signal to bounded, baseline-centered feedback
  update weights using centered target features, decay to w_0, and projection
end for

Variance control summary. Six mechanisms together keep the update stable in the experiments above: *(i)* bounded per-step feedback $p_t \in [-1, 1]$ via clipping or $\tanh$; *(ii)* EMA baseline $\bar{a}_t$ that removes the slow-moving component of p_t; *(iii)* clipped advantage $A_t \in [-1, 1]$; *(iv)* centered

feature delta $\Delta\phi_t = \phi_t(u_t^r) - \bar{\phi}_t^r$ that controls the bias of the update direction; *(v)* weight decay $\rho(w_0 - w_t)$ that anchors the linear weights to the interpretable heuristic initialization; and *(vi)* hard per-coordinate projection $\Pi_{[l,u]}$. Together they guarantee $\|w_{t+1} - w_t\|_\infty \leq \eta + \eta\rho\|u - l\|_\infty$ for every step regardless of feedback realization.

P.2 Per-Seed Scorer Results

Figure 18 and Table 16 report the seed-level spread behind the 21-run scorer ablation summarized in Section 4.4. They are intended as a reproducibility check on the mean and standard-error results in Table 3, rather than as a separate claim. The scatter plot shows how the reward-bandit, progress-based, hybrid, and adaptive variants vary across seeds, while the table lists the corresponding per-seed evaluation metrics used to compute the main ablation summary.

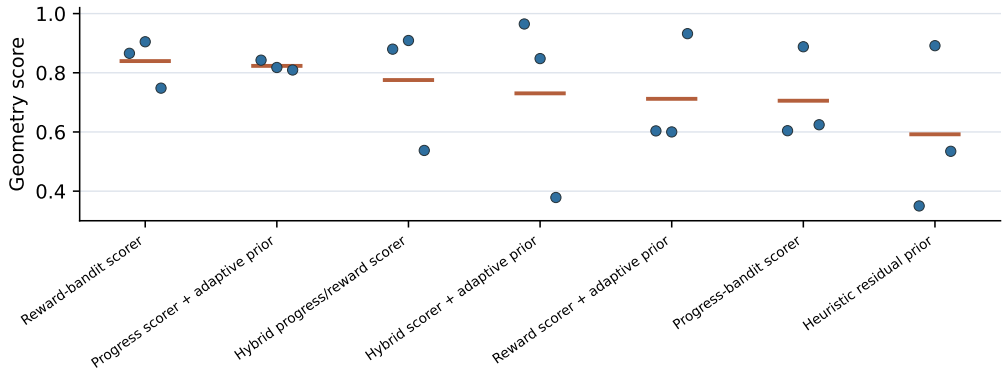


Figure 18: Per-seed geometry scores for the 21-run scorer ablation. Horizontal red segments denote condition means.

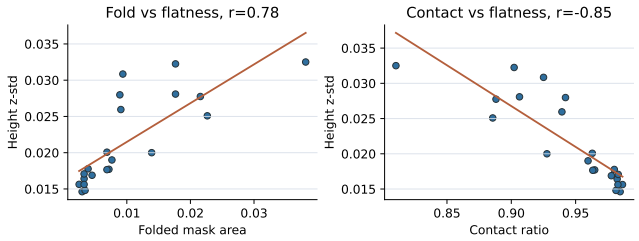


Figure 19: Fold, contact, and flatness across the 21-run scorer ablation: folded area versus height standard deviation ($r = 0.78$) and contact ratio versus height standard deviation ($r = -0.85$).

Table 16: Per-seed evaluation metrics for the scorer and adaptive-prior ablation.

Variant	Seed	Score	Return	Area	Fold ↓	Contact ↑
Reward-bandit scorer	940	0.866	30.6	0.207	0.0033	0.982
Reward-bandit scorer	941	0.905	32.7	0.213	0.0033	0.983
Reward-bandit scorer	942	0.748	37.0	0.220	0.0139	0.928
Progress scorer + adaptive prior	940	0.843	31.4	0.215	0.0072	0.965
Progress scorer + adaptive prior	941	0.818	31.3	0.210	0.0069	0.963
Progress scorer + adaptive prior	942	0.809	29.3	0.213	0.0077	0.960
Hybrid progress/reward scorer	940	0.880	32.2	0.214	0.0039	0.980
Hybrid progress/reward scorer	941	0.909	31.6	0.217	0.0033	0.983
Hybrid progress/reward scorer	942	0.538	17.4	0.179	0.0089	0.942
Hybrid scorer + adaptive prior	940	0.965	37.8	0.223	0.0030	0.985
Hybrid scorer + adaptive prior	941	0.848	37.0	0.220	0.0069	0.963
Hybrid scorer + adaptive prior	942	0.379	17.9	0.157	0.0094	0.925
Reward scorer + adaptive prior	940	0.604	27.5	0.190	0.0091	0.939
Reward scorer + adaptive prior	941	0.600	31.5	0.223	0.0227	0.886
Reward scorer + adaptive prior	942	0.932	37.8	0.225	0.0045	0.978
Progress-bandit scorer	940	0.604	30.3	0.227	0.0216	0.888
Progress-bandit scorer	941	0.888	29.4	0.208	0.0025	0.986
Progress-bandit scorer	942	0.624	28.0	0.220	0.0177	0.906
Heuristic residual prior	940	0.350	-68.0	0.229	0.0381	0.810
Heuristic residual prior	941	0.892	31.1	0.210	0.0034	0.981
Heuristic residual prior	942	0.535	23.0	0.210	0.0177	0.902

Q Additional Validation Across Initial States

The randomized campaign in Table 1 uses 3 seeds, 14 randomized training states, and 17 held-out physical drop-crumples. It extends the initial-state coverage of the fixed-state studies in Section 4 without replacing their results. The dual-affordance policy attains $F = .0066 \pm .001$ and $C = .968 \pm .007$, compared with $F = .0297 \pm .007$ and $C = .790 \pm .060$ for top-only SAC: 4.5× less fold and $+0.18$ contact. The displayed $\pm$ values denote standard errors. The fling-style and VLM-pull rows are the baselines described in Section 4.3.

Representation and sensing. The affordance channels inject a parameter-free geometric inductive bias; SAC still learns the action policy end to end. Under matched sensing, action space, reward, network, and training budget, replacing raw dual-view pixels with affordances improves F/C from $.0466/.766$ to $.0066/.968$: 7.1× less fold and $+.202$ contact. The matched comparison identifies the importance of a task-relevant state abstraction under limited robot interaction. The top-depth affordance condition further distinguishes the representation comparison from the additional contribution of bottom sensing (Table 17). These are comparisons within the randomized campaign; they should not be interpreted as changes to the separate long-horizon or scorer-ablation results.

Table 17: Gain decomposition in the randomized campaign. Relative folded-area changes and absolute contact changes are shown.

Comparison	Folded area F change	Contact C change
Top-only SAC → top-depth affordance	-70%	$+.167$
Top-depth affordance → dual affordance	-25%	$+.011$

R Sim-to-Real Transfer

Direct transfer and real-robot fine-tuning are evaluated across short-sleeve T-shirts, long-sleeve T-shirts, hoodies, pants, shorts, and towels. Table 18 reports the ranges of average folded area and contact. After 100 real-robot fine-tuning rollouts, both ranges shift toward improved unfolding. The policy input is the same six masks and robot state in both domains, with no raw depth image or point cloud. The simulated bottom mask thresholds the lowest cloth-particle height in each grid cell; on the real robot, bottom-view segmentation first determines the garment region seen through the glass, and bottom-camera depth then estimates the height of the visible cloth underside above the table plane in each cell, which is thresholded in the same way (Appendix G). The real estimate additionally carries depth-measurement, occlusion, and registration errors (Table 5), so the two masks correspond without being equivalent. These ranges summarize the transfer evaluation; they are distinct from the coverage gains of the physical comparison in Table 4.

Table 18: Ranges of average performance for direct sim-to-real transfer and after real-robot fine-tuning.

Transfer setting	Folded area $F \downarrow$	Contact $C \uparrow$
Direct sim-to-real	.035–.070	.74–.88
After 100 real-robot fine-tuning rollouts	.017–.055	.81–.94

S Camera Registration and Calibration

Support cues depend on spatial alignment between the two views. A double-sided calibration board reduces measured top–bottom registration error, restoring the benefit of complementary sensing (Table 19). As injected residual offset increases, folded area rises and contact falls smoothly. This registration evaluation complements the offline mask-shift probe and the synthetic policy-noise evaluation: each examines a different aspect of the same dependence on alignment.

Table 19: Registration quality and folded-area degradation before and after calibration. Degradation is measured relative to clean input.

Registration metric	Before	Calibrated
Median offset	6.0 cm (2.4 cells)	< 0.1 cell
Frames > 1 policy cell	94.1%	$< 3\%$
Mask IoU	.77	$\geq .97$
F degradation vs. clean	5–10$\times$	$\leq 1.1\times$

T Illumination Robustness

SAM 3 maintains stable cloth segmentation under moderate brightness changes, while very low light, side lighting, and overexposure degrade mask quality (Table 20). This result supports bounded tolerance to illumination variation, while retaining optical conditions as a deployment constraint. The $\pm$ values in this segmentation evaluation denote standard errors.

Table 20: SAM 3 garment-mask quality under illumination changes. Reported IoU uncertainties are standard errors. Relative drops are also shown.

Lighting	SAM 3 mask IoU	Drop
Standard	$.94 \pm .02$	—
Low light	$.91 \pm .03$	3.2%
Very low light	$.76 \pm .09$	19.1%
Side lighting	$.84 \pm .06$	10.6%
Overexposure	$.80 \pm .08$	14.9%

U Garment Coverage and Physical Evaluation

The garment set spans multiple categories and sizes, with fully unfolded areas listed in Table 21. In the physical comparison, each garment–method pair is evaluated over 11 trials. Coverage gain is final minus initial coverage, and the reported $\pm$ coverage summaries are mean$\pm$standard deviation. The raw dual-view coverage result is a range, $+.03$–$+.06$, rather than a mean with a dispersion estimate.

Table 4 reports both perception accuracy and manipulation outcomes. The paired dual-view versus top-only coverage difference is $+.06$, with 95% confidence interval $[+.02, +.10]$, alongside a 21% action reduction. These quantitative comparisons extend the physical evidence beyond the separately presented illustrative records and segmentation examples.

Table 21: Fully unfolded garment area (m^2) for four instances in each category.

Garment	G1	G2	G3	G4
Short-sleeve T-shirt	.36	.41	.47	.53
Long-sleeve T-shirt	.45	.52	.60	.68
Hoodie / sweatshirt	.52	.61	.71	.82
Pants / trousers	.35	.41	.48	.54
Shorts	.16	.20	.24	.29
Towel	.30	.48	.72	.98

Layering and cloth thickness. All evaluated methods fail on 3–4-layer stacks. Binary masks encode occupancy, not layer count, so complementary views do not eliminate the ambiguity of multilayer cloth. The fixed $\epsilon_{bot} = 3.5$ cm threshold nevertheless transfers to thin and thick analogs, with $F = .0080/.0088$, respectively. This thickness comparison does not remove the limitation on stacked layers.

Observability and hardware scope. Identical top-visible surfaces may conceal different folds and contact arrangements. For example, an inward fold of a layered skirt can remain hidden beneath a nearly unchanged top surface. Bottom sensing reduces this ambiguity but requires a transparent support: the reported setup uses an $80 glass table and $30 camera mounts. These are support and mounting costs, rather than the cost of the full robotic system. Future directions include using bottom-view information only as privileged supervision, or using top RGB-D with temporal history and active interaction at inference.

V Physical Evaluation Protocol and Per-State Results

The folded-area metric F measures cloth that is visible from above but not in contact with the table. It is therefore blind to two physical situations: a garment folded flat in half has both layers on the glass and yields $F \approx 0$ while remaining folded, and cloth hanging beyond the table edge leaves the bottom view without any contact to report. Coverage gain alone is also insufficient, because it depends on the initial crumple. To keep the physical evaluation from being certified by either quantity on its own, we evaluate by initial state and score the contact perception separately. Table 5 in the main text summarizes the results; Tables 22 and 23 state what each quantity tests.

Initial states and success. Trials are grouped by initial state: a flat double fold, a partial overhang that is released and allowed to settle before scoring, and an ordinary random crumple produced by the reset procedure of Appendix U. Final normalized coverage NC is the final cloth area divided by the fully unfolded area of the same garment (Table 21). A trial counts as a success only when it meets the NC standard of its state *and* passes an independent fold check by a human annotator using the top and bottom images; the normalized support residual $f_f = |M_{\text{fold}}|/|M_{\text{top}}|$ is reported as an auxiliary quantity. Each state uses 20 independent resets, and consecutive frames of a single recording never count as separate trials. Final NC is 0.90 for the flat double fold and the settled partial overhang and 0.85 for the random crumple, with success rates of 80%, 80%, and 70%; random crumples score lower because their difficulty varies more across resets. Timing runs from the first robot action; a trial fails if the criterion is not met within 720 s or if at least 50% of the garment area is dragged out of the top camera's view, and no separate cap on the number of actions is imposed. Final f_f is 0.05, and the standard deviation of final NC across trials is 0.08, so the effect is not carried by a few favorable trials. The physical comparison in Table 4 is a separate measurement: it uses coverage gain over 11 trials per garment–method pair and is not grouped by initial state.

Flat-folded initial states. A separate comparison starts from a flat-folded initial state: the garment appears flat but still contains folds or stacked layers, with the layers resting on the glass, so its initial support residual f_f is low and its near-table contact is high although it is not unfolded. It uses 3 garments held out from parameter tuning, with 20 independent resets per garment for each method (60 trials per method), and compares the full method (dual affordance) with two controls: the top-depth affordance control of Section 4.2, which computes the same affordance channels from top depth alone without the bottom view, and an edge-only control, which restricts pick targets to the silhouette-edge channel M_t^{edge} without the fold-contact frontier. Success follows the definition above: the state's NC standard and the independent fold check. The full method succeeds in 51 of 60 trials (85%), the top-depth affordance control in 36 of 60 (60%), and the edge-only control in 39 of 60 (65%): differences of 25 and 20 percentage points. Final normalized coverage NC is 0.91 for the full method and 0.84 for the edge-only control, an absolute difference of 0.07, so the full method not only passes the success criterion more often but also reaches higher final coverage; final NC was not recorded for the top-depth control (Table 5c and Table 22). Because the resets are independent, the methods do not share exactly matched initial configurations, and the comparison is reported at the level of aggregate success and mean final NC only. These 60 trials per method are a separate campaign from the 20 flat-double-fold resets above and are not pooled with them. The result shows that the full method still achieves real unfolding from a state whose support-related quantities already look favorable, and that both the bottom view and frontier targeting remain useful there; it does not make a low f_f or a high contact fraction a sufficient criterion for unfolding, and it does not show that bottom sensing alone resolves all layering ambiguity.

Perception quality. The action prior consumes the near-table (contact) estimate and its registration to the top view, so Table 23 reports the quality of that estimate over 60 independent test configura-

Table 22: Per-state physical unfolding results and what each quantity tests. NC is final normalized coverage; f_f is the normalized support residual.

Quantity	Value	What it tests
Flat double fold: final NC	0.90	Directly tests the "low F yet still folded" case: both layers touch the table, so $F \approx 0$ while the garment remains folded.
Partial overhang: final NC after release and settling	0.90	Excludes apparent gains while cloth hangs beyond the table edge and is invisible to the bottom view.
Random crumple: final NC	0.85	Difficulty varies more across resets; reported per state, never pooled.
Flat double fold: unfolding success rate	80%	Success requires the state's NC standard *and* an independent fold check.
Partial overhang: unfolding success rate	80%	Same definition.
Random crumple: unfolding success rate	70%	Covers distinct reset states.
Final normalized support residual f_f	0.05	Auxiliary only; cannot certify unfolding by itself.
Std. of final NC across trials	0.08	Shows the effect is not carried by a few favorable trials.
Independent trials per initial state	20	Independent resets; consecutive video frames are not trials.
Flat-folded state, held-out garments: full method success	51/60 (85%)	Three garments held out from parameter tuning, 20 independent resets per garment and method; success uses the same NC-plus-fold-check definition.
Flat-folded state, held-out garments: top-depth affordance success	36/60 (60%)	Same affordance channels from top depth alone, without the bottom view; the 25-percentage-point gap measures the bottom view's contribution in this state.
Flat-folded state, held-out garments: edge-only success	39/60 (65%)	Pick targets restricted to the silhouette-edge channel M_t^{edge} without the fold-contact frontier; the 20-percentage-point gap measures the contribution of frontier targeting.
Flat-folded state: final NC, full method / edge-only	0.91 / 0.84	Absolute difference of 0.07: the full method also reaches higher final coverage, not only more successes; not recorded for the top-depth control.

tions. Near-table mask quality is scored against annotated contact regions; the false near-table rate counts cloth that is physically above the ϵ_{bot} threshold yet labeled as contacting; the depth quantities report sensor validity and height error near the glass; and registration error is expressed in cells of the 32×32 policy grid (2.5 cm), the unit at which Appendix M shows a one-cell shift collapses the frontier. The contact-mask quality agrees with Table 4, and the registration error is consistent with the post-calibration values in Table 19.

Table 23: Near-table (contact) perception and top–bottom registration quality over 60 test configurations, and what each quantity tests.

Quantity	Value	What it tests
Near-table (contact) region IoU	0.91	Agreement with annotated contact regions.
Precision	0.93	Fraction of predicted contact that is annotated contact.
Recall	0.93	Fraction of annotated contact that is recovered.
Cloth above the height threshold ϵ_{bot} labeled near-table	10%	False near-table rate; cloth that is physically lifted yet labeled as contacting.
Valid-depth fraction inside the garment region	90%	Sensor validity near the glass.
Mean absolute height error w.r.t. the table plane	10 mm	Height accuracy near the glass.
95th percentile of absolute height error	20 mm	Tail of the height error.
Median top–bottom registration error	0.05 policy cell	Typical misalignment between the two views.
95th percentile registration error	1 policy cell	Tail misalignment; a one-cell shift collapses the frontier (Appendix M).
Independent test configurations	60	Distinct garment configurations scored.

W NOTATION

Table 24 summarizes the symbols used by the mask representation, residual controller, scorer, and evaluation metrics.

Table 24: Notation used in the method and evaluation.

Symbol	Meaning
s_t^p	Proprioceptive state at control step t.
M_t^{top}, M_t^{bot}	Top silhouette and bottom-contact masks.
$M_t^{fold}, M_t^{contact}$	Residual-fold and overlapping-contact channels.
$M_t^{frontier}, M_t^{edge}$	Fold-contact frontier and silhouette-edge channels.
$z_{table}, z_{i,j}^{min}$	Table height and lowest cloth surface in grid cell (i, j).
ϵ_{bot}	Near-table height threshold for the bottom mask.
$\phi_t(i, j), d^{out}$	Candidate-cell features and outward-distance feature.
$\mathcal{C}_t^r, u_t^r$	Feasible candidate set and selected cell for arm r.
$w_t, w_0, s_t(i, j)$	Scorer weights, heuristic initialization, and linear cell score.
$a_t^{RL}, a_t^{prior}, \lambda_t$	SAC action, prior action, and prior weight.
a_t^{exec}	Clipped residual action executed by the robot.
R_t, g_t, p_t	Environment reward, grounded geometry progress, and bounded feedback.
$\bar{a}_t, A_t, \Delta\phi_t$	EMA feedback baseline, clipped advantage, and centered feature delta.
$\eta, \rho, \beta, \epsilon$	Learning rate, decay, EMA coefficient, and exploration probability.
$\Pi_{[l,u]}$	Coordinate-wise projection onto the scorer weight bounds.
A, F, C, Z	Final area, folded mask area, contact ratio, and height standard deviation.
S	Geometry ranking index normalized within a table or figure.