RELIACT: POLICY-RELEVANT RELIABILITY FOR ROBUST UNDERWATER VISION-LANGUAGE-ACTION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Underwater vision-language-action (VLA) models encounter visual degradation and sensor failures in combinations absent from training. We establish an underwater VLA robustness benchmark on U0/USIM with 18 conditions and 3,560 closed-loop episodes, covering visual, sensor, and composite degradation while preserving dynamics and evaluation ground truth. Motivated by state-dependent policy impact and the limitations of fixed visual repair, we introduce ReliAct. It learns policy-relevant reliability from a frozen policy's response discrepancy under paired clean and corrupted observations. One predicted vector jointly regulates visual repair, multimodal routing, and action generation, without clean references at inference. ReliAct maintains clean performance and raises aggregate corrupted success from 26.7% to 36.0%. With single-source training, it improves success on three unseen composites by 17.1, 14.3, and 13.6 percentage points over U0. Adding 8.6M trainable parameters and 5.9% inference overhead, ReliAct establishes policy-relevant information use as an effective mechanism for robust underwater control across visual degradation, sensor failure, and unseen combinations.

1 INTRODUCTION

Vision-language-action (VLA) models transfer visual and linguistic knowledge to robot control (Brohan et al., 2023; Kim et al., 2024; Ghosh et al., 2024; Black et al., 2025; NVIDIA et al., 2025). USIM/U0 combines vision and onboard measurements for underwater navigation, grasping, transport, and tracking (Gu et al., 2025a). These observations face heterogeneous failures: absorption and scattering reduce visibility and distort color (Akkaynak & Treibitz, 2018), motion and occlusion obscure visual cues, and IMU bias or Doppler velocity log (DVL) outages compromise navigation measurements (Yampolsky et al., 2025). Some damaged visual content remains recoverable, whereas occlusion and stale measurements remove evidence. A robust policy must determine how much each source should influence its current decision.

Benchmarks expose VLA sensitivity to visual, sensor, and environmental perturbations (Fei et al., 2025; Jiang et al., 2026). Feature filtering and consistency training improve robustness (Fu et al., 2026; Guo et al., 2026), but fixed processing cannot capture every state-dependent effect. Our representative cases show different reference-policy response discrepancies across task phases under the same degradation; fixed visual repair also helps severe visual corruption much more than sensor failure. These observations motivate regulating information use according to its policy impact.

Figure 1(a) illustrates the inputs motivating *ReliAct*. Its policy-relevant reliability measures how well a corrupted source preserves a frozen policy's clean response. Paired observations supervise the ego camera, wrist camera, IMU, and DVL. At deployment, an estimator predicts their reliability from available observations and short temporal context, without clean references or corruption labels. One shared vector controls reliability-conditioned visual repair, reliability-guided routing, and reliability-conditioned action generation, connecting source-specific policy impact to information selection and control.

Figure 1(b) previews its gains across visual and sensor degradation. Our contributions are:

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

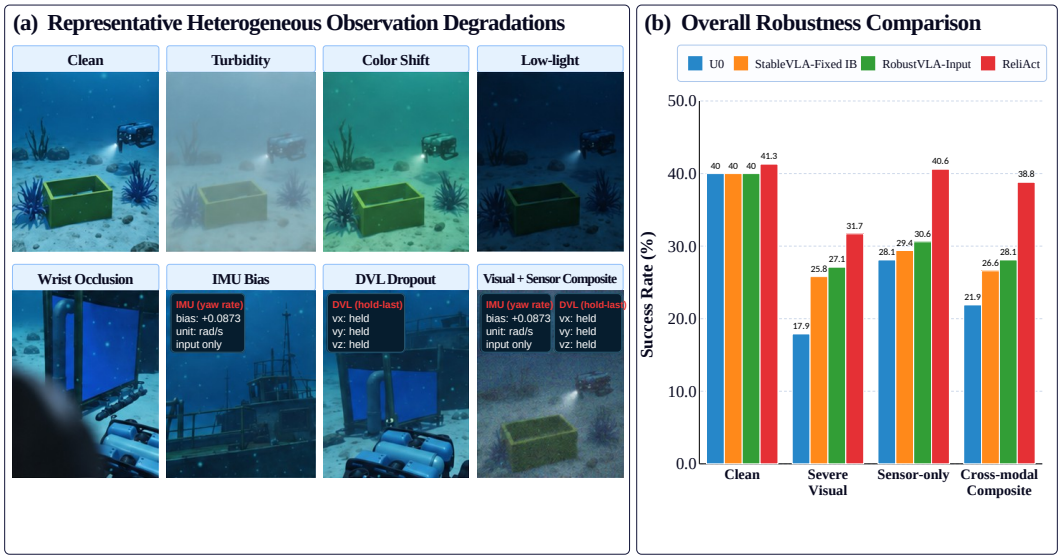

Figure 1: Representative observation degradations (a) and strong-baseline comparison (b). Severe visual denotes L3; cross-modal composite denotes visual–sensor degradation. Sensor annotations follow the input-only yaw-rate bias and hold-last protocols.

- **An underwater VLA robustness benchmark** covering 18 visual, sensor, and composite conditions, with paired evaluation and single-source training followed by unseen-composite testing.
- **Policy-relevant reliability** supervised by paired clean–corrupted policy responses, linking source reliability to task state and behavior.
- **Reliability-guided information use** jointly regulating visual repair, routing, and action generation with one predicted signal at low adaptation cost.

2 Related Work

Vision-language-action models. Generalist policies and diverse robot data broaden task and embodiment coverage (Ghosh et al., 2024; Open X-Embodiment Collaboration, 2024; Khazatsky et al., 2024), while diffusion and flow matching model continuous actions (Chi et al., 2023; Black et al., 2025). We study degraded observations in underwater control.

Robustness under observation degradation. Controlled evaluations reveal sensitivity to object, state, and observation changes (Fei et al., 2025; Zhou et al., 2025; Jiang et al., 2026). Domain randomization expands training coverage (Chen et al., 2025), and lifelong adaptation addresses changing environments (Wang et al., 2026b). Visual restoration reconstructs inputs (Guarnizo Orjuela et al., 2026); StableVLA filters nuisance features with an information bottleneck (IB) branch (Fu et al., 2026); RobustVLA combines input consistency with robustness to action perturbations (Guo et al., 2026); and STRONG-VLA separates robustness learning from task refinement (Xie et al., 2026). ReliAct adopts StableVLA's IB primitive, but learns source-specific policy-relevant reliability to control repair strength, multimodal routing, and action generation.

Reliability-aware fusion and decision-making. Evidence, uncertainty, and sensor availability guide multimodal fusion (Han et al., 2023; Paek & Kong, 2025; Wang et al., 2026a). Perceptual uncertainty informs navigation (Gao et al., 2026), while planning uncertainty can trigger help (Ren et al., 2023). VLA failure detection, success confidence, and action entropy support monitoring or observation reinjection (Gu et al., 2025b; Huang et al., 2026; Yang et al., 2026). ReliAct instead supervises sources through frozen-policy response discrepancies.

3 PRELIMINARY FINDINGS AND MOTIVATION

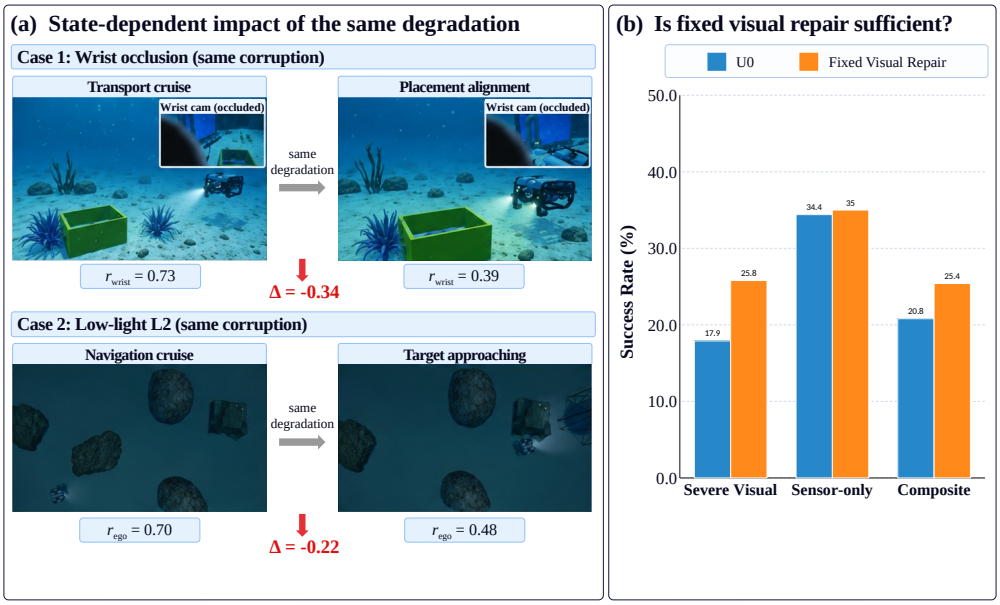

Figure 2: Motivation. (a) Representative phase-specific responses to unchanged degradation. (b) Fixed visual repair helps severe visual corruption but has limited benefit under sensor failure. Scores in (a) measure preservation of the reference policy's clean response.

Table 1: Policy-relevant reliability in representative task phases under the same degradation. Phase order follows the task sequence; Δr is the decrease between phases.

Degradation and phase transition	Earlier r	Later r	Δr
Wrist occlusion: grasp approach → final alignment	0.65	0.34	0.31
Wrist occlusion: transport cruise → placement alignment	0.73	0.39	0.34
Low-light L2: navigation cruise → target approach	0.70	0.48	0.22

Task-phase cases. U0 is vulnerable to both visual and sensor degradation; the consolidated results appear in Section 5.2. We examine whether degradation type and severity alone capture its policy impact. Using the paired-response score defined in Section 4.1, Table 1 and Figure 2(a) show lower reliability during alignment or target approach than during earlier phases under the same corruption. These cases motivate state-dependent information use: preserving a particular visual cue can matter more when the task requires precise localization.

Limits of fixed visual repair. On the ablation subset, fixed repair raises severe-visual success from 17.9% to 25.8%, but sensor-only success changes from 34.4% to 35.0%; composite success rises from 20.8% to 25.4% and remains limited (Figure 2(b)). Restoring representations therefore does not fully address missing or stale information. We use policy-relevant reliability to regulate repair, suppress unreliable sources, and expose the resulting information state to action generation.

4 METHOD

ReliAct adapts a frozen U0 backbone to heterogeneous observation degradation. Motivated by Section 3, it predicts a policy-relevant reliability vector and uses it throughout perception and action (Figure 3). The visual repair operator is inherited from StableVLA; the proposed mechanism is its joint regulation with routing and action generation through policy-derived reliability.

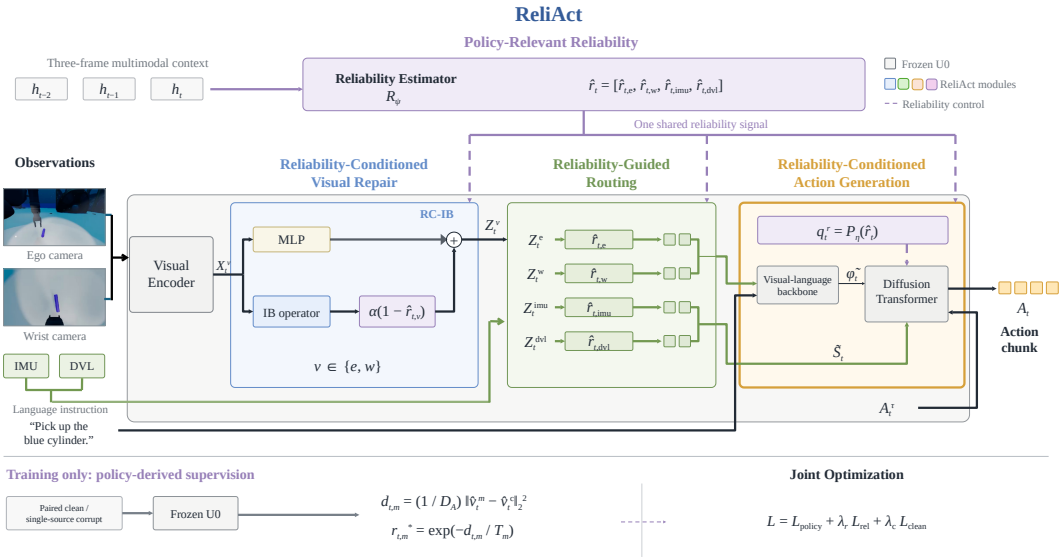

Figure 3: ReliAct. A three-frame estimator predicts ego-camera, wrist-camera, IMU, and DVL reliability. One shared vector regulates visual repair, source routing, and action generation. Paired clean–corrupted observations supply training supervision only; the U0 backbone remains frozen.

4.1 POLICY-RELEVANT RELIABILITY

Policy-derived target. U0 receives visual-language features ϕ_t and robot-state features S_t, and generates an action chunk with a diffusion Transformer (Gu et al., 2025a; Peebles & Xie, 2023). For action chunk A_t, noise ϵ, and flow-matching time τ (Lipman et al., 2023), the noisy action is

$$A_t^\tau = \tau A_t + (1-\tau)\epsilon. \tag{1}$$

Let $\mathcal{M} = \{e, w, \mathrm{imu}, \mathrm{dvl}\}$ denote the four observation sources. At the same state, we pair a clean observation o_t^c with o_t^m, which perturbs only source m. The other sources, language instruction, A_t, τ, and ϵ are shared. The frozen reference policy V_{θ_0} produces

$$\hat{v}_t^c = V_{\theta_0}(\phi_t^c, A_t^\tau, S_t^c), \qquad \hat{v}_t^m = V_{\theta_0}(\phi_t^m, A_t^\tau, S_t^m). \tag{2}$$

With D_A the number of scalar action entries, we measure the source-specific response discrepancy and map it to reliability:

$$d_{t,m} = \frac{1}{D_A}\|\hat{v}_t^m - \hat{v}_t^c\|_2^2, \tag{3}$$

$$r_{t,m}^* = \exp(-d_{t,m}/T_m), \qquad T_m > 0. \tag{4}$$

Larger $r_{t,m}^*$ means stronger preservation of the reference policy's clean response. This is a policy-relevant supervision target, not a measure of observation accuracy or task-success probability. It can vary with task state even when corruption type and severity are unchanged. The reference-policy dependence and interpretation limits are discussed in Appendix F.

Reliability estimation. Since Equation (4) requires a clean reference, it is used only for supervision. Let h_t denote multimodal features from the available observation at time t. A lightweight estimator uses three frames:

$$\hat{\boldsymbol{r}}_t = R_\psi(h_{t-2}, h_{t-1}, h_t) = [\hat{r}_{t,e}, \hat{r}_{t,w}, \hat{r}_{t,\mathrm{imu}}, \hat{r}_{t,\mathrm{dvl}}]. \tag{5}$$

The temporal context supplies short-term change cues relevant to motion blur and stale DVL measurements. At inference, the estimator receives no clean observation, corruption type, severity, or failure label.

4.2 Reliability-Guided Information Utilization

Reliability-Conditioned Visual Repair. For visual sources $v \in \{e, w\}$, we use the channel-covariance/IB operator of StableVLA (Fu et al., 2026) as a repair primitive. Its information-selection motivation relates to the information bottleneck (Tishby et al., 2000), and its channel-wise operation relates to cross-covariance attention (Ali et al., 2021). Given visual features X_t^v, reliability controls the residual strength:

$$Z_t^v = \mathrm{MLP}(X_t^v) + \alpha(1 - \hat{r}_{t,v})\,\mathrm{IB}(X_t^v), \qquad v \in \{e, w\}. \tag{6}$$

High-reliability inputs largely preserve the projected representation, whereas lower reliability increases repair. We call this reliability-conditioned IB repair (RC-IB).

Reliability-Guided Routing. Repair alone cannot restore absent evidence. We regulate the contribution of every source, including IMU and DVL:

$$\widetilde{Z}_t^m = \hat{r}_{t,m} Z_t^m + (1 - \hat{r}_{t,m}) N_m, \qquad m \in \mathcal{M}, \tag{7}$$

where N_m is a learnable neutral representation. Visual Z_t^m comes from Equation (6); sensor Z_t^m denotes its encoded features. Low reliability reduces a source's influence, consistent with robust fusion under missing modalities (Neverova et al., 2016; Choi & Lee, 2019). Repair and routing jointly allow continuous preservation, repair, and suppression without corruption-specific rules.

Reliability-Conditioned Action Generation. Per-source routing does not explicitly expose the full reliability state to the action module. We project the entire vector into an additional condition:

$$q_t^r = P_\eta(\hat{\boldsymbol{r}}_t), \tag{8}$$

$$\hat{v}_t^R = V_\Theta(\widetilde{\phi}_t, A_t^\tau, \widetilde{S}_t, q_t^r), \tag{9}$$

where $\widetilde{\phi}_t$ and $\widetilde{S}_t$ are constructed from routed features, and Θ comprises the frozen backbone and trainable lightweight adaptation. This condition lets action generation respond to the joint predicted reliability of the available sources.

4.3 Joint Optimization

Reliability estimation and policy adaptation are optimized jointly. The estimator learns the target in Equation (4) through

$$\mathcal{L}_{\mathrm{rel}} = \sum_{m \in \mathcal{M}} \mathrm{SmoothL1}(\hat{r}_{t,m}, r_{t,m}^*). \tag{10}$$

Policy consistency aligns ReliAct's corrupted response with the frozen clean response:

$$\mathcal{L}_{\mathrm{policy}} = \frac{1}{D_A} \left\| \hat{v}_t^R - \mathrm{sg}(\hat{v}_t^c) \right\|_2^2. \tag{11}$$

Gradients from $\mathcal{L}_{\mathrm{policy}}$ reach R_ψ through repair, routing, and action conditioning, coupling reliability to its downstream use. To preserve clean behavior, we additionally impose

$$\mathcal{L}_{\mathrm{clean}} = \frac{1}{D_A} \left\| \hat{v}_t^{R,c} - \mathrm{sg}(\hat{v}_t^c) \right\|_2^2 + \beta \left\| \hat{\boldsymbol{r}}_t^c - \mathbf{1} \right\|_2^2. \tag{12}$$

The full objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{policy}} + \lambda_r \mathcal{L}_{\mathrm{rel}} + \lambda_c \mathcal{L}_{\mathrm{clean}}. \tag{13}$$

Only lightweight adaptation components are trained. We generate paired samples from clean USIM trajectories using single-source degradations with continuously sampled strength. Fixed benchmark settings and all composite conditions are reserved for evaluation. At deployment, the clean branch and r^* are removed; ReliAct uses only observations and their three-frame context.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

Benchmark and tasks. We establish an underwater VLA robustness benchmark on U0/USIM (Gu et al., 2025a), retaining the original tasks and evaluator. Seven tasks cover navigation/inspection, grasping, transport, and tracking. The Core track evaluates all seven; Diagnostic and Stress tracks use task subsets appropriate to the affected modality. The full benchmark comprises 18 conditions and 3,560 closed-loop episodes per model (Appendix A).

Corruption and training protocols. The conditions comprise clean; nine Core visual corruptions (turbidity, color shift, and low-light at three levels); four Diagnostic conditions (overexposure, motion blur, wrist occlusion, and combined turbidity/low-light); and four Stress conditions (IMU bias, DVL dropout, and two visual–sensor combinations). Corruptions affect only model observations, with dynamics, ground truth, and the evaluator unchanged. Clean and corrupted episodes share task, seed, and initial state. This controlled design follows established corruption benchmarks (Hendrycks & Dietterich, 2019; Chattopadhyay et al., 2021; Kong et al., 2023; Xie et al., 2023); underwater visibility and color degradation are grounded in underwater imaging research (Akkaynak & Treibitz, 2018; Li et al., 2020; Akkaynak & Treibitz, 2019). We freeze the 3B U0 backbone and train on paired clean/single-source corrupted trajectories with continuous random strengths. All three composite conditions are unseen during training.

Baselines and evaluation subsets. We compare U0, U0 + CorrAug, StableVLA-Fixed IB, StableVLA-Fixed IB + SameAug, RobustVLA-Input, and Reliability-Gated U0. The StableVLA and RobustVLA baselines adapt their core mechanisms to U0 rather than using their original complete models (Fu et al., 2026; Guo et al., 2026). SameAug matches ReliAct's corruption sampling. Comparisons share a predefined evaluation manifest within each experiment. Section 5.2 uses the full benchmark; Sections 5.3–5.5 use separate fixed subsets for baselines, ablation, and reliability analysis. Absolute success rates across these subsets are not directly comparable. Appendix B details their scope.

5.2 OVERALL ROBUSTNESS RESULTS

Table 2: Overall robustness on the full benchmark. SR is success rate (%); Δ is the gain in percentage points computed before rounding. Composite rows are included in the track totals.

Evaluation / condition	N	U0	**ReliAct**	Δ
Clean	240	39.6	**40.8**	+1.3
Core (corrupted)	2,160	27.0	**35.5**	+8.5
Diagnostic	600	24.8	**35.7**	+10.8
Stress	560	27.9	**38.4**	+10.5
All corrupted	3,320	26.7	**36.0**	+9.2
Turbidity L2 + low-light L2	140	18.6	**35.7**	+17.1
Turbidity L2 + IMU bias	140	20.7	**35.0**	+14.3
Low-light L2 + DVL dropout	140	22.1	**35.7**	+13.6

ReliAct maintains clean performance (40.8% versus 39.6%) and improves all 17 corrupted conditions. Aggregate corrupted success increases from 888/3,320 to 1,195/3,320, a gain of 9.25 percentage points (Table 2). The reported paired-bootstrap 95% confidence interval is $[+7.7, +10.8]$ points, with McNemar $p < 10^{-30}$. Gains span Core, Diagnostic, and Stress tracks. Full counts and paired statistics appear in Appendix C.

As Figure 4 shows, ReliAct improves L3 turbidity, color shift, and low-light by 14.6, 10.4, and 15.0 points. Its benefit also extends beyond visual repair: IMU-bias success increases from 32.9% to 40.7%, and DVL-dropout success from 35.7% to 42.1%. Despite single-source training, gains on the three unseen composites reach 17.1, 14.3, and 13.6 points. These results support transferring reliability-guided information use to the tested combinations of damaged sources.

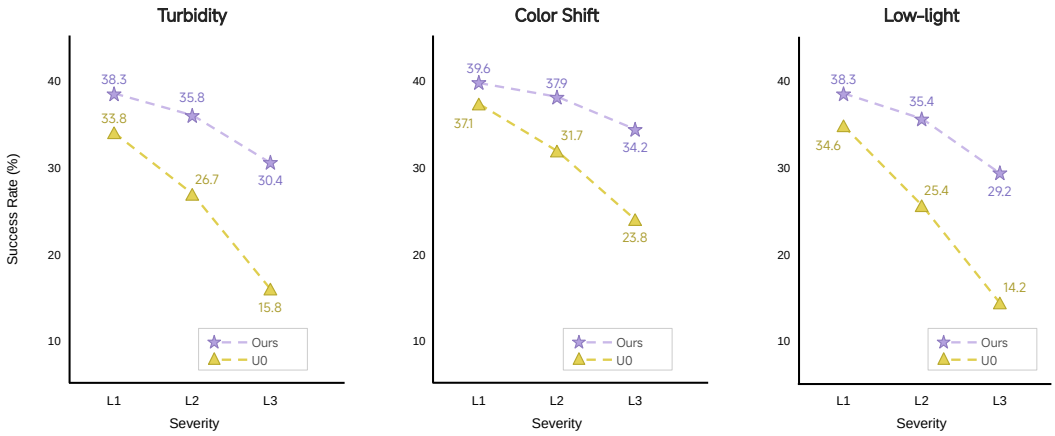

Figure 4: Robustness under visual degradation on the full Core track. ReliAct's advantage over U0 increases with severity across turbidity, color shift, and low-light. Each point covers 240 episodes.

5.3 Comparison with Strong Robustness Baselines

Table 3: Strong-baseline comparison on fixed evaluation subsets. All entries are SR (%). L1–L3 average the three Core visual families; visual diagnostic includes the visual composite. Sensor and cross-modal columns use a separate baseline subset.

Method	Clean	L1	L2	L3	Visual diag.	Sensor only	Cross-modal comp.
U0	40.0	35.4	27.9	17.9	25.7	28.1	21.9
U0 + CorrAug	38.8	37.9	31.7	22.9	29.6	30.0	25.3
StableVLA-Fixed IB	40.0	39.2	34.2	25.8	32.5	29.4	26.6
RobustVLA-Input	40.0	39.6	35.0	27.1	34.3	30.6	28.1
Fixed IB + SameAug	40.0	40.4	35.8	28.3	35.0	31.3	29.4
Reliability-Gated U0	40.0	38.3	33.8	24.6	31.1	36.6	32.8
ReliAct (ours)	**41.3**	**40.8**	**37.9**	**31.7**	**39.3**	**40.6**	**38.8**

Visual degradation. ReliAct achieves the highest success across visual severity levels among the evaluated adaptations (Table 3). It is close to StableVLA-Fixed IB + SameAug at L1, with a larger advantage at L2 (+2.1 points) and L3 (+3.4 points). At L3, ReliAct reaches 31.7%, compared with 28.3% for Fixed IB + SameAug and 27.1% for RobustVLA-Input, while maintaining clean performance.

Diagnostic and sensor degradation. ReliAct reaches 39.3% on visual diagnostic conditions, 4.3 points above Fixed IB + SameAug. On sensor-only and cross-modal composite conditions, it reaches 40.6% and 38.8%, exceeding the strongest baseline in these columns, Reliability-Gated U0, by 4.0 and 6.0 points. Relative to RobustVLA-Input, the corresponding gains are 10.0 and 10.7 points. Coordinating reliability with action generation and visual repair thus improves over the evaluated visual adaptations and direct gating baseline.

5.4 Ablation Studies

Incremental system gains. We progressively extend the adaptation on four representative tasks with 20 seeds per task (Table 4). Replacing fixed IB with RC-IB raises severe-visual success from 25.8% to 29.6%. Adding routing increases sensor-only success from 35.6% to 42.5%. Adding action conditioning then increases composite success from 32.5% to 37.5%. Clean performance is preserved. These are gains along an incremental configuration path; model capacity also changes, so they do not isolate each component's independent contribution.

Table 4: Incremental system gains on the same four-task manifest. Clean, severe visual, sensor-only, and composite groups contain 80, 240, 160, and 240 episodes, respectively. Entries are SR (%).

Variant	Clean	Severe visual	Sensor-only	Composite
U0	40.0	17.9	34.4	20.8
StableVLA-Fixed IB	40.0	25.8	35.0	25.4
RC-IB	41.3	29.6	35.6	28.3
RC-IB + Routing	41.3	30.0	42.5	32.5
ReliAct (full)	**41.3**	**31.7**	**43.8**	**37.5**

Joint optimization. Keeping architecture and capacity fixed, we compare a separately fitted, frozen estimator against joint optimization in which $\mathcal{L}_{\text{policy}}$ updates the estimator. Severe-visual, sensor-only, and composite success changes from 30.4/40.6/33.8% to 31.7/43.8/37.5%, respectively (Appendix D.1). This supports coupling reliability estimation to downstream adaptation. ReliAct trains 8.6M additional parameters, approximately 0.29% of U0's 3B backbone. Mean latency increases from 0.940 to 0.995 s (+5.9%); complete efficiency measurements appear in Appendix E.3.

5.5 Policy-Relevant Reliability Analysis

Prediction agreement. On approximately 20,000 held-out temporal windows, the estimator achieves a macro MAE of 0.068 and Spearman correlation of 0.785 against the policy-derived target (Table 5). Prediction uses only available observations and three-frame context. These results measure agreement with r^*; its closed-loop utility is examined separately below.

Table 5: Agreement with the policy-derived reliability target on held-out windows. Lower MAE/RMSE and higher Spearman correlation are better.

Observation source	MAE ↓	RMSE ↓	Spearman ρ ↑
Ego camera	0.071	0.104	0.780
Wrist camera	0.079	0.113	0.740
IMU	0.063	0.095	0.800
DVL	0.060	0.091	0.820
Macro average	0.068	0.101	0.785

Closed-loop value of the signal. We fix all parameters of the trained full model and replace only its reliability input on a 280-episode subset. Predicted reliability achieves 40.4% success, above All-one (32.9%) and Random (27.9%; Table 6). Thus, the predicted signal has decision value within the trained framework beyond the presence of additional model parameters. An Oracle target computed from paired clean–corrupted observations reaches 46.4%, providing a privileged reference unavailable at deployment.

Table 6: Closed-loop reliability controls on 280 episodes. The four full-model rows share one trained checkpoint; U0 and Fixed IB are separate references. Oracle uses additional clean information.

Model / reliability input	Successes	SR (%) ↑	Policy distance ↓
U0	82	29.3	0.067
StableVLA-Fixed IB	96	34.3	0.050
Full model: Random	78	27.9	0.071
Full model: All-one	92	32.9	0.054
ReliAct: predicted $\hat{\boldsymbol{r}}$	**113**	**40.4**	**0.034**
Full model: Oracle target $\boldsymbol{r}^*$	130	46.4	0.024

Unseen corruption families. We additionally evaluate leave-one-corruption-out (LOCO) training, withholding motion blur or DVL dropout from training. LOCO reaches 42.5% on motion blur (U0: 34.4%; full ReliAct: 45.0%) and 40.7% on DVL dropout (U0: 35.7%; full: 42.1%). Defining gain retention as $(\mathrm{SR}_{\mathrm{LOCO}} - \mathrm{SR}_{\mathrm{U0}})/(\mathrm{SR}_{\mathrm{full}} - \mathrm{SR}_{\mathrm{U0}})$ gives 0.76 and 0.78, respectively. Retaining most of the gain on these withheld families supports generalization beyond the corruption types encountered in training. Appendix D.2 reports the complete comparison.

Auxiliary diagnostic. Predicted reliability lowers policy distance to 0.034, versus 0.054 for All-one and 0.071 for Random. Because this metric uses the same discrepancy family as r^*, it measures target consistency rather than independent validation of the reliability definition.

6 Discussion and Limitations

The results connect observation degradation to how a policy uses information. Low predicted reliability increases visual repair while reducing the source's routed weight; as reliability approaches zero, routing approaches a neutral representation. Action conditioning exposes the full information state to control. This offers a mechanism for the observed gains under sensor failure and unseen combinations, where reconstructing visual features alone is insufficient.

Policy-relevant reliability is defined with respect to a frozen reference policy: a high score need not imply accurate sensing or a correct action. Single-source targets do not explicitly model nonlinear interactions between simultaneous failures, and routing cannot replace evidence when several critical sources are unavailable. Our evaluation covers U0/USIM simulation and three unseen composites; broader combinations, other backbones, and real underwater deployment remain open. Appendix F discusses these boundaries and the interpretation of the diagnostic analyses.

7 Conclusion

We establish an underwater VLA robustness benchmark and introduce ReliAct, which learns policy-relevant reliability from paired-policy response discrepancies. A shared predicted vector jointly regulates visual repair, multimodal routing, and action generation. ReliAct maintains clean performance while improving robustness to visual degradation, sensor failure, and three unseen composite conditions, with 8.6M additional trainable parameters. These results support policy-relevant information use as an effective basis for robust underwater VLA control.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing, literature discovery, drafting and restructuring selected passages, methodological and experimental discussion, figure preparation, and generation or modification of software code used in implementation and analysis. All AI-assisted code was reviewed, tested, and validated by the authors before use. The authors verified all citations, experimental settings, results, analyses, and scientific claims and take full responsibility for the final manuscript. AI tools were not used to autonomously generate reported results, synthetic evaluation datasets, or mathematical proofs.

REPRODUCIBILITY STATEMENT

Implementation details needed to reproduce ReliAct and the benchmark are provided in Appendix G, including architecture, reliability supervision, training hyperparameters, corruption sampling, evaluation manifests, and paired statistical tests. Appendices A–C specify the benchmark, evaluation subsets, and condition-level results; the U0 backbone is frozen unless otherwise stated.

REFERENCES

D. Akkaynak and T. Treibitz. A Revised Underwater Image Formation Model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

D. Akkaynak and T. Treibitz. Sea-Thru: A Method for Removing Water From Underwater Images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Alaaeldin Ali, Hugo Touvron, Mathilde Caron, Piotr Bojanowski, Matthijs Douze, Armand Joulin, Ivan Laptev, Natalia Neverova, Gabriel Synnaeve, Jakob J. Verbeek, and Hervé Jégou. XCiT: Cross-Covariance Image Transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Robert Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, Laura Smith, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang, and Ury Zhilinsky. π_0: A Vision-Language-Action Flow Model for General Robot Control. In *Robotics: Science and Systems (RSS)*, 2025. URL https://arxiv.org/abs/2410.24164.

Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, Pete Florence, Chuyuan Fu, Montse Gonzalez Arenas, Keerthana Gopalakrishnan, Kehang Han, Karol Hausman, Alexander Herzog, Jasmine Hsu, Brian Ichter, Alex Irpan, Nikhil Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Isabel Leal, Lisa Lee, Tsang-Wei Edward Lee, Sergey Levine, Yao Lu, Henryk Michalewski, Igor Mordatch, Karl Pertsch, Kanishka Rao, Krista Reymann, Michael Ryoo, Grecia Salazar, Pannag Sanketi, Pierre Sermanet, Jaspiar Singh, Anikait Singh, Radu Soricut, Huong Tran, Vincent Vanhoucke, Quan Vuong, Ayzaan Wahid, Stefan Welker, Paul Wohlhart, Jialin Wu, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Tianhe Yu, and Brianna Zitkovich. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. In *Conference on Robot Learning (CoRL)*, 2023. URL https://arxiv.org/abs/2307.15818.

P. Chattopadhyay, J. Hoffman, R. Mottaghi, and A. Kembhavi. RobustNav: Towards Benchmarking Robustness in Embodied Navigation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

Tianxing Chen, Zanxin Chen, Baijun Chen, Zijian Cai, Yibin Liu, Zixuan Li, Qiwei Liang, Xianliang Lin, Yiheng Ge, Zhenyu Gu, Weiliang Deng, Yubin Guo, Tian Nian, Xuanbing Xie, Qiangyu Chen, Kailun Su, Tianling Xu, Guodong Liu, Mengkang Hu, Huan-ang Gao, Kaixuan Wang, Zhixuan Liang, Yusen Qin, Xiaokang Yang, Ping Luo, and Yao Mu. RoboTwin 2.0: A Scalable Data Generator and Benchmark with Strong Domain Randomization for Robust Bimanual Robotic Manipulation. *arXiv preprint arXiv:2506.18088*, 2025. URL https://arxiv.org/abs/2506.18088.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. In *Robotics: Science and Systems (RSS)*, 2023. URL https://arxiv.org/abs/2303.04137.

J.-H. Choi and J.-S. Lee. EmbraceNet: A Robust Deep Learning Architecture for Multimodal Classification. *Information Fusion*, 51:259–270, 2019.

Senyu Fei, Siyin Wang, Junhao Shi, Zihao Dai, Jikun Cai, Pengfang Qian, Li Ji, Xinzhe He, Shiduo Zhang, Zhaoye Fei, Jinlan Fu, Jingjing Gong, and Xipeng Qiu. LIBERO-Plus: In-depth Robustness Analysis of Vision-Language-Action Models. *arXiv preprint arXiv:2510.13626*, 2025. URL https://arxiv.org/abs/2510.13626.

Yiyang Fu, Chubin Zhang, Shukai Gong, Yufan Deng, Kaiwei Sun, Qiyang Min, Qibin Hou, Yansong Tang, Jianan Wang, and Daquan Zhou. StableVLA: Towards Robust Vision-Language-Action Models without Extra Data. In *International Conference on Machine Learning (ICML)*, 2026. URL https://arxiv.org/abs/2605.18287.

Jianzhe Gao, Rui Liu, Yuxuan Xu, Tongtong Cao, Yingxue Zhang, Zhanguang Zhang, Sida Peng, Yi Yang, and Wenguan Wang. Uncertainty-Aware Gaussian Map for Vision-Language Navigation. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://arxiv.org/abs/2605.26503.

Dibya Ghosh, Homer Rich Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, Jianlan Luo, You Liang Tan, Lawrence Yunliang Chen, Quan Vuong, Ted Xiao, Pannag R Sanketi, Dorsa Sadigh, Chelsea Finn, and Sergey Levine. Octo: An Open-Source Generalist Robot Policy. In *Robotics: Science and Systems (RSS)*, 2024. URL https://arxiv.org/abs/2405.12213.

Junwen Gu, Zhiheng Wu, Pengxuan Si, Shuang Qiu, Zhentao Zhang, Yukai Feng, Luoyang Sun, Laien Luo, Lianyi Yu, Jian Wang, and Zhengxing Wu. USIM and U0: A Vision-Language-Action Dataset and Model for General Underwater Robots. *arXiv preprint arXiv:2510.07869*, 2025a. URL https://arxiv.org/abs/2510.07869.

Qiao Gu, Yuanliang Ju, Shengxiang Sun, Igor Gilitschenski, Haruki Nishimura, Masha Itkina, and Florian Shkurti. SAFE: Multitask Failure Detection for Vision-Language-Action Models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025b. URL https://arxiv.org/abs/2506.09937.

Daniel Yezid Guarnizo Orjuela, Leonardo Scappatura, Veronica Di Gennaro, Riccardo Andrea Izzo, Gianluca Bardaro, and Matteo Matteucci. Improving Robustness of Vision-Language-Action Models by Restoring Corrupted Visual Inputs. *arXiv preprint arXiv:2602.01158*, 2026. URL https://arxiv.org/abs/2602.01158.

Jianing Guo, Zhenhong Wu, Chang Tu, Yiyao Ma, Xiangqi Kong, Zhiqian Liu, Jiaming Ji, Shuning Zhang, Yuanpei Chen, Kai Chen, Qi Dou, Yaodong Yang, Xianglong Liu, Huijie Zhao, Weifeng Lv, and Simin Li. RobustVLA: On Robustness of Vision-Language-Action Model against Multi-Modal Perturbations. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://arxiv.org/abs/2510.00037.

Z. Han, C. Zhang, H. Fu, and J. T. Zhou. Trusted Multi-View Classification with Dynamic Evidential Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2551–2566, 2023.

D. Hendrycks and T. Dietterich. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. In *International Conference on Learning Representations (ICLR)*, 2019. URL https://arxiv.org/abs/1903.12261.

Dehao Huang, Aoxiang Gu, Chengjie Zhang, Bolin Zou, Wenlong Dong, Zilang Cen, Yue Wang, and Hong Zhang. VLAConf: Calibrated Task-Success Confidence for Vision-Language-Action Models. *arXiv preprint arXiv:2605.29605*, 2026. URL https://arxiv.org/abs/2605.29605.

Huaide Jiang, Yash Chaudhary, Yuping Wang, Zehao Wang, Raghav Sharma, Manan Mehta, Yang Zhou, Lichao Sun, Zhiwen Fan, Zhengzhong Tu, and Jiachen Li. NavTrust: Benchmarking Trustworthiness for Embodied Navigation. *arXiv preprint arXiv:2603.19229*, 2026. URL https://arxiv.org/abs/2603.19229.

Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti, Soroush Nasiriany, Mohan Kumar Srirama, Lawrence Yunliang Chen, Kirsty Ellis, Peter David Fagan, Joey Hejna, Masha Itkina, Marion Lepert, Yecheng Jason Ma, Patrick Tree Miller, Jimmy Wu, Suneel Belkhale, Shivin Dass, Huy Ha, Arhan Jain, Abraham Lee, Youngwoon Lee, Marius Memmel, Sungjae Park, Ilija Radosavovic, Kaiyuan Wang, Albert Zhan, Kevin Black, Cheng Chi, Kyle Beltran Hatch, Shan Lin, Jingpei Lu, Jean Mercat, Abdul Rehman, Pannag R Sanketi, Archit Sharma, Cody Simpson, Quan Vuong, Homer Rich Walke, Blake Wulfe, Ted Xiao, Jonathan Heewon Yang, Arefeh Yavary, Tony Z. Zhao, Christopher Agia, Rohan Baijal, Mateo Guaman Castro, Daphne Chen, Qiuyu Chen, Trinity Chung, Jaimyn Drake, Ethan Paul Foster, Jensen Gao, Vitor Guizilini, David Antonio Herrera, Minho Heo, Kyle Hsu, Jiaheng Hu, Muhammad Zubair Irshad, Donovon Jackson, Charlotte Le, Yunshuang Li, Kevin Lin, Roy Lin, Zehan Ma, Abhiram Maddukuri, Suvir Mirchandani, Daniel Morton, Tony Nguyen, Abigail O'Neill, Rosario Scalise, Derick Seale, Victor Son, Stephen Tian, Emi Tran, Andrew E. Wang, Yilin Wu, Annie Xie, Jingyun Yang, Patrick Yin, Yunchu Zhang, Osbert Bastani, Glen Berseth,

Jeannette Bohg, Ken Goldberg, Abhinav Gupta, Abhishek Gupta, Dinesh Jayaraman, Joseph J Lim, Jitendra Malik, Roberto Martín-Martín, Subramanian Ramamoorthy, Dorsa Sadigh, Shuran Song, Jiajun Wu, Michael C. Yip, Yuke Zhu, Thomas Kollar, Sergey Levine, and Chelsea Finn. DROID: A Large-Scale In-The-Wild Robot Manipulation Dataset. In *Robotics: Science and Systems (RSS)*, 2024.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. OpenVLA: An Open-Source Vision-Language-Action Model. In *Conference on Robot Learning (CoRL)*, 2024. URL https://arxiv.org/abs/2406.09246.

L. Kong, S. Xie, H. Hu, L. X. Ng, B. R. Cottereau, and W. T. Ooi. RoboDepth: Robust Out-of-Distribution Depth Estimation under Corruptions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao. An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020.

Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le. Flow Matching for Generative Modeling. In *International Conference on Learning Representations (ICLR)*, 2023.

N. Neverova, C. Wolf, G. W. Taylor, and F. Nebout. ModDrop: Adaptive Multi-Modal Gesture Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1692–1706, 2016.

NVIDIA, Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Jim Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, Joel Jang, Zhenyu Jiang, Jan Kautz, Kaushil Kundalia, Lawrence Lao, Zhiqi Li, Zongyu Lin, Kevin Lin, Guilin Liu, Edith Llontop, Loic Magne, Ajay Mandlekar, Avnish Narayan, Soroush Nasiriany, Scott Reed, You Liang Tan, Guanzhi Wang, Zu Wang, Jing Wang, Qi Wang, Jiannan Xiang, Yuqi Xie, Yinzhen Xu, Zhenjia Xu, Seonghyeon Ye, Zhiding Yu, Ao Zhang, Hao Zhang, Yizhou Zhao, Ruijie Zheng, and Yuke Zhu. GR00T N1: An Open Foundation Model for Generalist Humanoid Robots. *arXiv preprint arXiv:2503.14734*, 2025. URL https://arxiv.org/abs/2503.14734.

Open X-Embodiment Collaboration. Open X-Embodiment: Robotic Learning Datasets and RT-X Models. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

Dong-Hee Paek and Seung-Hyun Kong. Availability-aware Sensor Fusion via Unified Canonical Space. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. URL https://arxiv.org/abs/2503.07029.

W. Peebles and S. Xie. Scalable Diffusion Models with Transformers. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Allen Z. Ren, Anushri Dixit, Alexandra Bodrova, Sumeet Singh, Stephen Tu, Noah Brown, Peng Xu, Leila Takayama, Fei Xia, Jake Varley, Zhenjia Xu, Dorsa Sadigh, Andy Zeng, and Anirudha Majumdar. Robots That Ask For Help: Uncertainty Alignment for Large Language Model Planners. In *Conference on Robot Learning (CoRL)*, 2023.

N. Tishby, F. C. Pereira, and W. Bialek. The Information Bottleneck Method. *arXiv preprint arXiv:physics/0004057*, 2000. URL https://arxiv.org/abs/physics/0004057.

Junyao Wang, Yulin Xu, Yu Li, Pramod Khargonekar, and Mohammad Abdullah Al Faruque. CRUISE: Vision-Language Model-Guided Uncertainty-Aware Cross-Modal Sensor Fusion for Robust Autonomous Driving. *arXiv preprint arXiv:2608.09202*, 2026a. URL https://arxiv.org/abs/2608.09202.

Xudong Wang, Gan Li, Zhiyu Liu, Yao Wang, Lianqing Liu, and Zhi Han. All-day Multi-scenes Lifelong Vision-and-Language Navigation with Tucker Adaptation. In *International Conference on Learning Representations (ICLR)*, 2026b. URL https://arxiv.org/abs/2603.14276.

S. Xie, L. Kong, W. Zhang, J. Ren, L. Pan, K. Chen, and Z. Liu. RoboBEV: Towards Robust Bird's Eye View Perception under Corruptions. *arXiv preprint arXiv:2304.06719*, 2023. URL https://arxiv.org/abs/2304.06719.

Yuhan Xie, Yuping Yan, Yunqi Zhao, Handing Wang, and Yaochu Jin. STRONG-VLA: Decoupled Robustness Learning for Vision-Language-Action Models under Multimodal Perturbations. *arXiv preprint arXiv:2604.10055*, 2026. URL https://arxiv.org/abs/2604.10055.

Z. Yampolsky, N. Cohen, and I. Klein. Transformer-Based Robust Underwater Inertial Navigation in Prolonged Doppler Velocity Log Outages. In *OCEANS*, 2025. URL https://arxiv.org/abs/2504.07697.

Jiabing Yang, Yixiang Chen, Yuan Xu, Peiyan Li, Xiangnan Wu, Zichen Wen, Bowen Fang, Tao Yu, Zhengbo Zhang, Yingda Li, Kai Wang, Jing Liu, Nianfeng Liu, Yan Huang, and Liang Wang. UAOR: Uncertainty-aware Observation Reinjection for Vision-Language-Action Models. *arXiv preprint arXiv:2602.18020*, 2026. URL https://arxiv.org/abs/2602.18020.

Xueyang Zhou, Yangming Xu, Guiyao Tie, Yongchao Chen, Guowen Zhang, Duanfeng Chu, Pan Zhou, and Lichao Sun. LIBERO-PRO: Towards Robust and Fair Evaluation of Vision-Language-Action Models Beyond Memorization. *arXiv preprint arXiv:2510.03827*, 2025. URL https://arxiv.org/abs/2510.03827.

A BENCHMARK SPECIFICATION

A.1 TASK COVERAGE AND OBSERVATION ISOLATION

The benchmark uses the original U0/USIM task definitions and success evaluator. Table 7 lists the seven tasks and their Core-track episode counts. The clean condition and each of the nine Core corruptions cover all seven tasks, giving $10 \times 240 = 2,400$ episodes. Diagnostic conditions account for 600 episodes and Stress conditions for 560, totaling 3,560 episodes per model, including 3,320 corrupted episodes.

Table 7: Tasks and full-Core coverage. Seeds are 0–39 for 40-episode tasks and 0–19 for 20-episode tasks.

ID	Task	Capability	N/condition
G	`goto_water_tower`	Navigation	40
S	`scan_ship_modern`	Inspection	20
R	`pick_red_shallow`	Grasping	40
P	`pick_pipe0_shallow`	Grasping	40
B	`pick_blue_factory`	Grasping	40
T	`transfer_red_shallow`	Transport	40
F	`follow_boat`	Tracking	20
	Total		240

Perturbations are applied along the model observation path, after sensor/image generation and before policy input construction. Simulator pose, target state, task goals, true dynamics, and success thresholds are unchanged. In particular, IMU bias changes only the yaw angular-velocity observation, not the orientation angle or physical robot state. DVL dropout keeps the last valid velocity; it does not substitute zeros or NaNs or change the message frequency. The sensor annotations in Figure 1(a) summarize these input effects rather than independent measured trajectories.

A.2 CONDITION REGISTRY

Table 8: The 18 benchmark conditions. Task IDs are defined in Table 7; "all" denotes all seven tasks. L1–L3 entries each represent three separate conditions.

Condition	Fixed evaluation setting	Tasks	N each
Clean	No corruption	All	240
Turbidity L1–L3	Transmission $t = 0.8, 0.6, 0.4$	All	240
Color shift L1–L3	Red-channel gain $0.75, 0.50, 0.25$	All	240
Low-light L1–L3	Brightness $0.7, 0.4, 0.2$, plus noise	All	240
Overexposure	Local 15% area, episode-static	G,R,T,F	140
Motion blur	$\lvert$yaw rate$\rvert > 0.027$ rad/s; 11-pixel kernel	G,S,R,T,F	160
Wrist occlusion	Wrist camera only; 15% static area	R,P,B,T	160
Turbidity + low-light	Both at L2	G,R,T,F	140
IMU yaw bias	Yaw angular velocity $+0.0873$ rad/s	G,R,T,F	140
DVL dropout	0.5 s every 10 s; hold-last	G,R,T,F	140
Turbidity + IMU	Turbidity L2 and IMU yaw bias	G,R,T,F	140
Low-light + DVL	Low-light L2 and DVL dropout	G,R,T,F	140

Visual severity is fixed within each episode. Low-light noise may change between frames, with randomness derived from the episode seed. Overexposure and occlusion use episode-static masks. Motion blur is triggered by the yaw-rate threshold. The benchmark protocol requires fixed composition order and recording of corruption settings, masks, blur triggers, and DVL dropout intervals. Wrist-only occlusion is evaluated on tasks that use the wrist camera rather than forcing it onto tasks without that input.

A.3 Pairing, metrics, and aggregation

Paired runs share the task, seed, initial robot pose and orientation, object and target placement, instruction, waypoints, and environment randomization. An initial-state manifest records the reset state when a seed alone is insufficient for deterministic replay. Only the policy-side corruption differs between conditions. Task-aware tracks use the matched task set for their clean denominator.

Success rate is the number of successful episodes divided by the number of valid episodes. For condition c evaluated on task set $\mathcal{T}_c$, retention is

$$\text{Retention}(c) = \frac{\text{SR}(\mathcal{T}_c, c)}{\text{SR}(\mathcal{T}_c, \text{clean})}. \tag{14}$$

Raw retention is not clipped to $[0, 1]$ and can exceed one. A zero clean denominator makes the ratio undefined. Track success aggregates episode counts; the PRS-style score averages matched-condition retention. Gains in percentage points are calculated before display rounding, so subtracting two one-decimal printed rates can differ from the reported gain by 0.1 points.

The original task metrics are retained: success weighted by path length (SPL) for navigation, average successful duration (ASD) for grasping, stage success rate (SSR) for transport, and mean target distance (MTD) for tracking. ASD averages successful episodes within each task; the grasp-family ASD is a macro-average over the three grasping tasks.

B Evaluation Subsets and Training Scope

B.1 Comparison manifests

The full benchmark, strong-baseline comparison, ablation, and reliability-control study have different evaluation scopes. Table 9 makes these distinctions explicit. Every comparison within a study shares its evaluation set; its absolute SR should not be compared with a rate from a different study. In particular, the sensor/composite subset in Table 3 differs from the ablation subset in Table 4, explaining their different U0 sensor rates.

Table 9: Scope of the evaluation sets used in the paper. Counts apply per method or reliability setting.

Study	Scope	Episode count
Full benchmark	Seven tasks; Core plus task-aware Diagnostic and Stress	3,560 total; 3,320 corrupted
Strong baselines	Four canonical tasks (G,R,T,F), 20 seeds; clean and visual conditions	Approximately 1,080 visual episodes
Baseline sensors	Separate fixed sensor/composite subset; distinct from ablation manifest	Independent of the visual count
Incremental ablation	Four canonical tasks, 20 seeds per task	80 clean; 240 severe visual; 160 sensor; 240 composite
Joint optimization	Same canonical subset as incremental ablation	Same group counts
Reliability prediction	Held-out temporal windows; offline	Approximately 20,000 windows
Reliability controls	One fixed canonical subset; shared full-model checkpoint	280 episodes
LOCO	Motion blur or DVL dropout, on their full-benchmark task subsets	160 or 140 episodes

B.2 Training and baseline definitions

Paired samples are constructed from clean USIM trajectories. Training uses a single degradation at a time with continuous random strength; no composite condition is used in training. Fixed benchmark severities are evaluation settings, rather than a set of discrete condition labels supplied to the estimator. Except in LOCO, individual degradation families can be present in training even though their evaluated combinations are unseen. This setting tests composition transfer, not universal novelty of every elementary corruption.

Source-specific targets compare a clean observation against a version in which source m alone is corrupted; the other sources remain clean. The teacher shares action chunk, flow time, and noise across the two branches. ReliAct jointly optimizes the estimator, repair, routing, and action adaptation while freezing the U0 backbone. The clean branch is used only for supervision and consistency. At inference, the estimator has no clean reference, condition ID, severity label, DVL drop flag, or blur-trigger flag.

U0 + CorrAug applies corruption augmentation. StableVLA-Fixed IB adapts the StableVLA-style visual branch to U0 with fixed repair strength. StableVLA-Fixed IB + SameAug uses the same corruption sampling as ReliAct. RobustVLA-Input adapts the input-consistency mechanism to the same backbone; it is not the original complete RobustVLA system, including all output-side robustness components. Reliability-Gated U0 uses reliability directly for feature gating. Matching the backbone and, for SameAug, corruption sampling controls these factors; it does not imply that every baseline has the same trainable parameter count or an identical complete optimization procedure.

B.3 Interpreting supervision and controls

The reliability-control experiment changes the signal in one already-trained model. It establishes that this model uses its predicted reliability effectively relative to constant or random input. It does not compare independently retrained versions supervised by alternative definitions such as sensor-health labels, image quality, entropy, or task-success probability. Likewise, the incremental ablation changes both functionality and trainable capacity. The same-capacity Separate-versus-Joint comparison specifically tests optimization coupling, while the fixed-checkpoint controls test signal substitution.

C Complete Robustness Results

C.1 Clean task performance

Table 10 reports the seven-task clean results. For context, the original U0 paper reports 43.1% overall SR under its 20-task/700-trial protocol (Gu et al., 2025a). The 39.6% seven-task reference corresponds to the selected tasks and is not the original paper's overall rate. The selected-task values in the source results also match the corresponding published U0 task values.

Table 10: Clean task results. Parentheses show successful episodes; task IDs follow Table 7.

Task	N	U0 SR	**ReliAct** SR	Metric	U0	**ReliAct**
G	40	75.0 (30)	**77.5 (31)**	SPL	0.822	0.829
S	20	85.0 (17)	**85.0 (17)**	SPL	0.647	0.655
R	40	20.0 (8)	**22.5 (9)**	ASD (s)	48.81	47.2
P	40	22.5 (9)	**22.5 (9)**	ASD (s)	79.53	76.0
B	40	32.5 (13)	**35.0 (14)**	ASD (s)	113.35	107.5
T	40	25.0 (10)	**25.0 (10)**	SSR (%)	45.0	47.5
F	20	40.0 (8)	**40.0 (8)**	MTD (m)	3.61	3.48
Overall	240	39.6 (95)	**40.8 (98)**			

Table 11: Clean performance by task family on the seven-task selection. Each cell reports U0 / ReliAct.

Capability	SR (%)	Original task metric
Navigation/inspection	78.3 / **80.0**	SPL: 0.764 / 0.776
Grasping	25.0 / **26.7**	ASD: 80.6 / 76.9 s
Transport	25.0 / **25.0**	SSR: 45.0 / 47.5%
Tracking	40.0 / **40.0**	MTD: 3.61 / 3.48 m

C.2 ALL CONDITIONS AND PAIRED STATISTICS

Table 12 consolidates the complete condition-level counts and rates, including the representative U0 degradation results used to motivate the method. Table 13 reports the track-level retention summaries. Composite rows are subsets of the Diagnostic and Stress tracks, not extra episodes.

Table 12: Complete 18-condition results. SR is in percent; retention uses the matched clean task set.

		U0		**ReliAct**		Retention	
Condition	N	Succ.	SR	Succ.	SR	U0	**ReliAct**
Clean	240	95	39.6	98	**40.8**	1.000	1.000
Turbidity L1	240	81	33.8	92	**38.3**	0.853	0.939
Turbidity L2	240	64	26.7	86	**35.8**	0.674	0.878
Turbidity L3	240	38	15.8	73	**30.4**	0.400	0.745
Color shift L1	240	89	37.1	95	**39.6**	0.937	0.969
Color shift L2	240	76	31.7	91	**37.9**	0.800	0.929
Color shift L3	240	57	23.8	82	**34.2**	0.600	0.837
Low-light L1	240	83	34.6	92	**38.3**	0.874	0.939
Low-light L2	240	61	25.4	85	**35.4**	0.642	0.867
Low-light L3	240	34	14.2	70	**29.2**	0.358	0.714
Overexposure	140	43	30.7	57	**40.7**	0.768	0.983
Motion blur	160	55	34.4	72	**45.0**	0.753	0.960
Wrist occlusion	160	25	15.6	35	**21.9**	0.625	0.833
Turbidity L2 + low-light L2	140	26	18.6	50	**35.7**	0.464	0.862
IMU yaw bias	140	46	32.9	57	**40.7**	0.821	0.983
DVL dropout	140	50	35.7	59	**42.1**	0.893	1.017
Turbidity L2 + IMU	140	29	20.7	49	**35.0**	0.518	0.845
Low-light L2 + DVL	140	31	22.1	50	**35.7**	0.554	0.862

Table 13: Track-level PRS-style retention. These averages use matched task sets within each condition and need not equal the ratio of track-level aggregate SRs.

Method	Core	Diagnostic	Stress
U0	0.682	0.653	0.696
ReliAct	**0.868**	**0.910**	**0.927**

Across all corrupted episodes, the paired outcome counts are 435 U0-failure/ReliAct-success pairs and 128 U0-success/ReliAct-failure pairs. These imply 760 shared successes and 1,997 shared failures. Thus $\Delta\,\mathrm{SR} = (435 - 128)/3,320 = 9.24699$ percentage points. The reported paired-bootstrap 95% interval is $[7.7, 10.8]$ points and the reported McNemar result is $p < 10^{-30}$. The confidence interval summarizes the reported aggregate paired analysis; it is not a new estimate of seed-clustered uncertainty across repeated conditions. The marginal counts alone do not reconstruct individual episode or task correlations.

D ADDITIONAL RELIABILITY AND ADAPTATION ANALYSES

D.1 JOINT VERSUS SEPARATE OPTIMIZATION

Separate training fits the reliability target and then freezes the estimator during downstream adaptation. Joint training allows the policy-consistency loss to update the estimator. Architecture and capacity are held fixed. The results in Table 14 support joint optimization along this comparison; the incremental architecture changes are reported separately in Table 4.

D.2 LEAVE-ONE-CORRUPTION-OUT GENERALIZATION

LOCO-Visual withholds motion blur from training and evaluates motion blur; LOCO-Sensor withholds DVL dropout and evaluates DVL dropout. The full model uses single-source training, including the constituent families but no composite samples. Consequently, the three composite conditions are

Table 14: Joint optimization ablation. All entries are SR (%) except the gain row, which is in percentage points. Gains use unrounded rates.

Training	Severe visual	Sensor-only	Composite	All corrupted
Separate	30.4	40.6	33.8	34.9
Joint (**ReliAct**)	**31.7**	**43.8**	**37.5**	**37.7**
Gain	+1.3	+3.1	+3.8	+2.8

unseen combinations for the full model itself. These are complementary generalization tests: LOCO withholds an elementary family, whereas composite testing withholds the joint occurrence of known elementary degradations.

Table 15: Generalization to withheld families and unseen combinations. All SRs are percentages. Full denotes standard single-source ReliAct training.

Test condition	N	U0	Full	LOCO / unseen	Gain retention
Motion blur	160	34.4	45.0	**42.5**	0.76
DVL dropout	140	35.7	42.1	**40.7**	0.78
Turbidity L2 + low-light L2	140	18.6	35.7	**35.7**	—
Turbidity L2 + IMU	140	20.7	35.0	**35.0**	—
Low-light L2 + DVL	140	22.1	35.7	**35.7**	—

Gain retention uses unrounded success rates. On motion blur, the counts implied by the reported rates are 55 U0 successes, 72 full-model successes, and 68 LOCO successes, giving $(68-55)/(72-55) = 0.7647$. On DVL dropout, 50, 59, and 57 successes give $(57-50)/(59-50) = 0.7778$. The retained gains demonstrate transfer in these two withheld-family tests. They do not establish generalization to every unseen sensor failure or arbitrary joint failure.

D.3 Additional estimator behavior

Table 16 reports mean predicted reliability under clean and corrupted observations, complementing the target-agreement errors in Table 5. Severity and temporal summaries appear in Table 17. These are behavioral diagnostics, rather than alternative ground-truth reliability labels.

Table 16: Mean predicted reliability by observation source.

Source	Clean	Corrupted
Ego camera	0.95	0.50
Wrist camera	0.94	0.45
IMU	0.97	0.61
DVL	0.98	0.56
Macro average	0.96	0.53

The reported temporal delay spans failure onset to the reliability crossing criterion in logged observations. Two control frames at 10 Hz correspond to 0.20 s, which should not be confused with the approximately one-second policy-inference latency. The phase-specific cases in Table 1 likewise illustrate different responses under the same degradation; they are not a controlled causal study of task phase.

Table 17: Reliability sensitivity to severity and transient failure. The reported monotonicity summary is descriptive. Detection delay is measured offline on the 10 Hz observation stream, not as end-to-end VLA reaction latency.

Visual corruption	L1	L2	L3	Monotonicity
Turbidity	0.84	0.61	0.34	84%
Color shift	0.91	0.76	0.51	79%
Low-light	0.82	0.57	0.29	86%
Mean	0.857	0.647	0.380	83%
Temporal failure	**Normal**	**Failure**	**Decrease**	**Detection delay**
Motion blur	0.90	0.62	0.28	0.20 s
DVL hold-last	0.97	0.58	0.39	0.20 s

E Offline Diagnostics and Efficiency

E.1 Action prediction and representation consistency

Offline diagnostics use a held-out seven-task paired clean/corrupted subset. Table 18 reports action prediction error (APE), feature consistency, and policy-flow recovery. APE is tied to this subset and should not be compared directly with the original U0 paper's headline offline aggregate of 0.0359. Feature consistency is the cosine similarity of matched clean/corrupted features. Policy-flow distance uses shared flow time and noise; recovery is $1 - D_{\mathrm{ReliAct}}/D_{\mathrm{U0}}$.

Table 18: Offline action and representation diagnostics. Lower APE is better; higher feature consistency and policy-flow recovery are better.

Method	Clean APE	Visual APE	Sensor APE	Composite APE
U0	0.059	0.132	0.095	0.162
CorrAug	0.061	0.107	0.088	0.139
Fixed IB	0.060	0.092	0.093	0.125
ReliAct	**0.058**	**0.075**	**0.067**	**0.087**

Feature stage	U0	Fixed IB	**ReliAct**
Vision encoder	0.68	0.68	0.68
Projector / repair	0.44	0.66	**0.75**
DiT conditioning	0.27	0.58	**0.79**

Degradation group	U0 flow distance	**ReliAct** flow distance	Recovery
Mild	0.033	0.025	24.2%
Moderate	0.064	0.039	39.1%
Severe visual	0.103	0.051	50.5%
Composite	0.128	0.052	59.4%

The policy-distance control in Table 6 and flow recovery here both use the same velocity-field discrepancy family as the reliability target. They therefore indicate consistency with the construction, rather than independent evidence that this definition is superior to alternative reliability measures. Closed-loop success provides the separate downstream outcome.

E.2 Auxiliary target-error audit

Table 19 preserves the offline CAP target-error audit. Because establishing a mechanism through CAP requires verifying that repaired features actually enter the CAP path, these values are not used as evidence for the proposed adaptation mechanism. Metrics retain the task-family conventions of the original audit and are not compared across families.

Table 19: Auxiliary CAP target-error audit. These values are excluded from the main mechanism claims.

Task family	Clean U0	Corrupted U0	Corrupted **ReliAct**	Reduction
Navigation	1.40	2.22	1.48	33.3%
Grasping	0.032	0.056	0.036	35.7%
Transport	0.056	0.091	0.059	35.2%
Tracking	0.124	0.194	0.130	33.0%

E.3 Parameter and inference overhead

The efficiency protocol uses the same hardware and batch size one for all models, at least 1,000 timed inferences, and an initial 100-inference warm-up. The models retain U0's 10 Hz control frequency, 1.6 s inference interval, 16-step action horizon, and four Euler steps. Table 20 reports mean and 95th-percentile latency along with peak GPU memory. Control frequency describes action execution, not how often a full VLA forward pass is run.

Table 20: Efficiency with a frozen 3B U0 backbone. Added and trainable parameters are in millions, memory in GB, and latency in seconds.

Method	Added	Trainable	Peak mem.	Mean	P95	Overhead
U0	0	—	21.0	0.940	1.090	—
StableVLA-Fixed IB	4.8	4.8	21.4	0.968	1.125	+3.0%
Reliability-Gated U0	2.6	2.6	21.3	0.966	1.121	+2.8%
ReliAct	**8.6**	**8.6**	21.9	0.995	1.162	+5.9%

The additional 8.6M parameters are approximately $8.6/3,000 = 0.287\%$ of the backbone. Mean overhead is $(0.995/0.940 - 1) \times 100\% = 5.85\%$, rounded to 5.9%. These measured parameter and timing quantities support the lightweight adaptation claim; no FLOP count is inferred from latency.

F Interpretation and Scope

Reference-policy dependence. The reliability target quantifies response preservation for a particular frozen policy, state, action sample, flow time, and discrepancy scale. A source that the reference policy ignores can receive a high target score even when it is inaccurate. Preserving the clean response can also preserve the reference policy's errors. Changing the reference backbone therefore requires re-evaluating both the target and its estimator. Policy-relevant reliability should not be interpreted as objective sensor health or a calibrated probability that an action will succeed.

Single-source targets under composite degradation. Each training target isolates a source while other sources remain clean. Under a composite test condition, ReliAct predicts the complete vector from the jointly corrupted observations; it does not obtain clean interventions at deployment. The training targets do not explicitly decompose nonlinear interactions among simultaneously corrupted sources. The three unseen-composite results demonstrate empirical transfer to those conditions, but do not prove that the model has learned a general law of compositional failure. Similarly, the two LOCO experiments provide evidence on motion blur and DVL dropout, rather than arbitrary unseen families.

Repair and suppression. Combining Equations (6) and (7) clarifies the visual mechanism. A low source score increases the repair residual before routing, but its routed contribution is still multiplied by reliability. As the score approaches zero, the output approaches the neutral representation. Thus, stronger repair does not force reliance on fully failed content. Action conditioning additionally supplies the complete predicted source state; the incremental ablation is consistent with its utility on composites, without isolating a causal contribution from capacity changes.

Evidence limits. The phase examples are representative cases rather than independently validated universal phase rules. Final alignment may depend more on local cues than cruising, but the present

comparisons do not separately manipulate every possible state-dependent factor. The Oracle target uses clean information and is a privileged diagnostic, not guaranteed ground-truth reliability or a formal upper bound on task performance. Replacing predicted input with a constant or random signal can create a distribution shift relative to training; the controls demonstrate usefulness within this model, not optimality among all possible reliability definitions.

Unavailable information and deployment. Routing reduces interference from an unreliable source but cannot reconstruct evidence that is absent from every available source. Several simultaneous critical failures, long outages beyond the temporal context, or fundamentally ambiguous observations can leave the policy without enough information to choose a correct action. Our experiments are confined to U0/USIM simulation. Transfer to real underwater platforms, other VLA backbones, arbitrary combinations, and prolonged complete sensor loss remains to be established.

G Reproducibility Statement

To facilitate reproduction of ReliAct and the underwater VLA robustness benchmark established in this work, we provide additional implementation details on the architecture, reliability supervision, training hyperparameters, corruption sampling, evaluation manifests, and statistical tests. Unless otherwise stated, all experiments use the same configuration. The 3B-parameter U0 backbone remains frozen, and only the added ReliAct modules are optimized.

G.1 ReliAct Implementation Details

Reliability estimator. The reliability estimator R_ψ takes multimodal features from three consecutive control timesteps, (h_{t-2}, h_{t-1}, h_t), and predicts reliability for the ego camera, wrist camera, IMU, and DVL:

$$\hat{\boldsymbol{r}}_t = R_\psi(h_{t-2}, h_{t-1}, h_t) = [\hat{r}_{t,e}, \hat{r}_{t,w}, \hat{r}_{t,\mathrm{imu}}, \hat{r}_{t,\mathrm{dvl}}] \in [0,1]^4. \tag{15}$$

U0's observation/control stream updates at 10 Hz, so consecutive control observations are separated by 0.1 s, and three timesteps span 0.2 s of recent history. This captures transient changes such as motion blur and stale DVL measurements without requiring corruption IDs, severity labels, dropout flags, or blur triggers.

The 0.1 s interval refers to the observation/control stream, rather than full VLA policy inference. The system maintains a rolling buffer of recent control observations. At each policy inference, it selects observations from the latest three control timesteps $(t-2, t-1, t)$ and obtains their representations through the frozen U0 feature interface for R_ψ. Thus, the estimator uses the most recent 0.2 s of history, while the full U0 diffusion policy retains a 1.6 s inference interval and generates a 16-step action chunk. The 3B-parameter U0 policy need not run at 10 Hz.

We reuse the existing multimodal feature interface of frozen U0 to represent each control observation as a 512-dimensional fused feature. This feature projection belongs to the frozen backbone and adds no trainable parameters. Concatenating the three features produces a 1,536-dimensional input to a three-layer MLP:

$$1536 \longrightarrow 1024 \longrightarrow 512 \longrightarrow 4. \tag{16}$$

The first two layers use GELU activations and dropout of 0.1; the output layer uses a sigmoid. This MLP constitutes the trainable reliability estimator R_ψ, with approximately 2.1M parameters. ReliAct adds 8.6M trainable parameters in total, as detailed in Table 21.

Action-side adaptation uses rank-8 LoRA only in the final two conditioning/cross-attention blocks of the diffusion Transformer. All other U0 parameters remain frozen. The added parameters constitute approximately 0.29% of the 3B-parameter U0 backbone, consistent with the lightweight adaptation described in the main text.

G.2 Reliability Targets and Key Hyperparameters

For each observation source $m \in \{e, w, \mathrm{imu}, \mathrm{dvl}\}$, training pairs a clean observation with one in which only source m is corrupted. With the same robot state, action chunk, flow-matching time τ,

Table 21: Trainable parameters added by ReliAct.

Module	Trainable parameters
Reliability estimator R_ψ	2.1M
RC-IB visual repair	4.8M
Neutral representations and routing	0.1M
Reliability projection P_η	0.1M
Action-side LoRA	1.5M
Total	8.6M

and diffusion noise ϵ, frozen U0 produces two velocity-field responses. Their discrepancy is

$$d_{t,m} = \frac{1}{D_A} \|\hat{v}_t^m - \hat{v}_t^c\|_2^2, \tag{17}$$

and the corresponding policy-relevant reliability target is

$$r_{t,m}^* = \exp(-d_{t,m}/T_m). \tag{18}$$

The four observation sources use separate, fixed temperatures:

$$T_e = 0.08, \qquad T_w = 0.08, \qquad T_{\text{imu}} = 0.06, \qquad T_{\text{dvl}} = 0.06. \tag{19}$$

These values are set solely from the scale of policy-response discrepancies in the training data and fixed before formal benchmark evaluation. The same values are used for all main experiments, ablations, and composite evaluations, without adjustment for individual test conditions.

Reliability-conditioned visual repair uses

$$Z_t^v = \text{MLP}(X_t^v) + \alpha(1 - \hat{r}_{t,v})\,\text{IB}(X_t^v), \tag{20}$$

with a fixed repair coefficient $\alpha = 0.3$. The joint training objective is

$$\mathcal{L} = \mathcal{L}_{\text{policy}} + \lambda_r \mathcal{L}_{\text{rel}} + \lambda_c \mathcal{L}_{\text{clean}}, \tag{21}$$

where $\lambda_r = 1.0$ and $\lambda_c = 0.5$. Section 4.3 defines the clean-preservation loss and its reliability regularizer, for which $\beta = 0.1$. All formal experiments therefore use $(\alpha, \beta, \lambda_r, \lambda_c) = (0.3, 0.1, 1.0, 0.5)$.

G.3 Training Settings

ReliAct is initialized from a pretrained U0 checkpoint, with the entire 3B-parameter backbone frozen. Training uses clean USIM trajectories and generates paired clean/single-source-corrupted observations online. Each training sample perturbs only one observation source, so no composite degradation is used during training. The three composite conditions are reserved for testing.

The flow-matching time and diffusion noise are sampled as

$$\tau \sim \mathcal{U}(0, 1), \qquad \epsilon \sim \mathcal{N}(0, I). \tag{22}$$

The clean and corrupted forward passes used to generate reliability targets share exactly the same action chunk, τ, and ϵ, so their response discrepancy arises only from the perturbed observation source. Table 22 summarizes the implementation and training hyperparameters.

Training uses AdamW with a 500-step linear warm-up followed by cosine learning-rate decay. All main results use the same hyperparameters, without adjusting the training configuration for specific test degradations.

Single-source corruption sampling. To avoid reducing the estimator to a classifier of fixed severity levels, training does not use the benchmark's discrete L1/L2/L3 labels. Instead, strengths are sampled continuously within each degradation family, with the discrete motion-blur kernel choices listed in Table 23.

Only one degradation is sampled at a time, and composite corruptions never appear as training inputs. LOCO-Visual and LOCO-Sensor additionally exclude the entire motion-blur and DVL-dropout families, respectively.

Table 22: ReliAct implementation and training hyperparameters.

Setting	Value
Base VLA	U0, 3B
Frozen backbone	Yes
Trainable parameters	8.6M
Estimator history	3 frames
Frame interval	0.1 s
Temporal span	0.2 s
Estimator hidden dimensions	1024 / 512
Estimator dropout	0.1
$T_e/T_w/T_{\mathrm{imu}}/T_{\mathrm{dvl}}$	0.08 / 0.08 / 0.06 / 0.06
RC-IB coefficient α	0.3
Clean regularization β	0.1
λ_r/λ_c	1.0 / 0.5
Flow time τ	$\mathcal{U}(0,1)$
Action-side LoRA rank	8
Optimizer	AdamW
Learning rate	2×10^{-4}
Weight decay	0.01
Adam (β_1, β_2)	(0.9, 0.999)
Global batch size	64
Training steps	20,000
Warm-up	500 steps
Learning-rate schedule	Cosine decay
Gradient clipping	1.0
Precision	BF16
Training seed	0
Control frequency	10 Hz
Inference interval	1.6 s
Action horizon	16
Euler steps	4

Table 23: Single-source corruption sampling during training.

Training degradation	Sampling
Turbidity	Transmission $t \sim \mathcal{U}(0.4, 0.8)$
Color shift	Red-channel gain $R \sim \mathcal{U}(0.25, 0.75)$
Low-light	Brightness $\sim \mathcal{U}(0.2, 0.7)$, plus noise
Local overexposure	Area ratio $\sim \mathcal{U}(0.10, 0.20)$
Motion blur	Kernel $\in \{7, 9, 11\}$ pixels
Wrist occlusion	Area ratio $\sim \mathcal{U}(0.10, 0.20)$
IMU yaw bias	Magnitude $\sim \mathcal{U}(0.03, 0.0873)$ rad/s, random sign
DVL dropout	Duration $\sim \mathcal{U}(0.3, 0.7)$ s, hold-last

G.4 Benchmark Protocol and Evaluation Manifests

Appendix A specifies the full benchmark. Appendix A.1 and Table 7 list the seven official tasks, task categories, episode counts, and seed ranges. Appendix A.2 and Table 8 define all 18 evaluation conditions and their exact parameters. Appendix A.3 describes paired evaluation, metric aggregation, and the initial-state replay protocol.

The benchmark comprises one clean condition, nine Core visual conditions, four Diagnostic conditions, and four Stress conditions: 18 conditions and 3,560 closed-loop episodes per model. Corruptions are applied only along the model observation path. Robot dynamics, ground-truth pose, target state, and task goals in Stonefish, together with the original U0 evaluator, remain unchanged. Clean and corrupted episodes share the same task, seed, and initial state.

Core visual parameters are turbidity transmission $t \in \{0.8, 0.6, 0.4\}$, color-shift red-channel gain $R \in \{0.75, 0.50, 0.25\}$, and low-light brightness in $\{0.7, 0.4, 0.2\}$. Diagnostic and Stress conditions additionally include 15% local overexposure, 11-pixel motion blur, 15% wrist occlusion, an IMU

yaw-rate bias of $+0.0873$ rad/s, and 0.5 s DVL hold-last dropout every 10 s, together with the corresponding composite conditions.

Experiments use separate evaluation manifests fixed in advance, with their scopes documented in Appendix B.1 and Table 9. The full benchmark, strong-baseline comparison, incremental ablation, joint-versus-separate optimization study, reliability-control study, and LOCO evaluations each have a fixed manifest. Within each experiment, all methods strictly share tasks, seeds, and initial states. Absolute SR values from different manifests are not compared directly.

Each manifest entry records at least the task ID, seed, initial robot state, target/object state, instruction, waypoints, environment seed, and corruption configuration. For degradations involving additional randomness, it also records overexposure/occlusion masks, motion-blur triggers, low-light noise seeds, corruption composition order, and DVL dropout intervals to support deterministic replay.

G.5 Statistical Tests and Code Release

Complete condition-level results and paired statistical analysis are provided in Appendix C.2. For individual-condition success rates, we report Wilson 95% confidence intervals. Overall comparisons between U0 and ReliAct preserve the pairing of episodes by task, seed, and initial state.

The 95% confidence interval for the success-rate difference is computed using 10,000 paired-bootstrap replicates. Each replicate resamples matched episode pairs with replacement within each condition, then aggregates using the original episode counts to obtain

$$\Delta\mathrm{SR} = \mathrm{SR}_{\mathrm{ReliAct}} - \mathrm{SR}_{\mathrm{U0}}. \tag{23}$$

The bootstrap random seed is fixed to 0. Binary success outcomes are also evaluated with a two-sided exact McNemar test. On the full corrupted benchmark,

$$n_{01} = 435, \qquad n_{10} = 128, \tag{24}$$

where n_{01} counts U0 failures paired with ReliAct successes, and n_{10} counts U0 successes paired with ReliAct failures. These yield an aggregate gain of approximately 9.25 percentage points, a paired-bootstrap 95% confidence interval of $[+7.7, +10.8]$ points, and McNemar $p < 10^{-30}$.

Upon publication, we will release the complete code and reproduction resources corresponding to the reported results. These include benchmark corruption hooks and condition configurations; training and inference configuration files; the ReliAct architecture and all hyperparameters; teacher-target generation scripts; task/seed/initial-state evaluation manifests; episode aggregation and retention calculation scripts; and the paired-bootstrap, Wilson-interval, and McNemar-test scripts used to produce the paper's statistical results.

The released manifests and configurations will correspond directly to the tasks, condition IDs, corruption parameters, and statistical conventions reported in Appendices A–C, enabling reconstruction of every evaluation subset and independent verification of the main results.