DUET: DUAL UNCERTAINTY EVIDENCE TRACKING WITH VISUAL TOKENS AND ACTION TRACES IN VLA

Anonymous authors
Paper under double-blind review

ABSTRACT

A frozen vision-language-action (VLA) policy may encounter unfamiliar visual content or multiple valid action continuations during deployment. We associate these cases with different uncertainty effects: coverage-related epistemic effects for unfamiliar visual content and action-related aleatoric effects for multiple valid continuations. Existing monitors often merge both into one score and obscure the source of a warning. DUET tracks the two evidence sources separately: a visual evidence branch measures novelty relative to successful demonstrations, while an action evidence branch measures local conditional dispersion at a fixed observation and noisy flow state. Separate references calibrate the two branches into EXECUTE, CAUTION, and HELP, and the visual branch also provides spatial evidence. The visual evidence branch detects unfamiliar observations with 0.940 frame AUROC and identifies the corresponding regions. The action evidence branch detects local action ambiguity with AUROC 0.963 ± 0.052. On a physical robot, HELP triggers in $45/50$ unfamiliar-object trials and $1/50$ familiar-object trials, with every unfamiliar alert preceding irreversible motion. DUET shows how a frozen VLA can track the provenance of uncertainty-related evidence and convert complementary evidence into calibrated deployment responses.

1 INTRODUCTION

A deployment monitor for a frozen vision–language–action (VLA) policy should preserve the provenance of its evidence. An observation may be unusual relative to successful demonstrations, producing a visual novelty signal. The same observation may admit several action continuations once a noisy flow state is fixed, producing a local action-dispersion signal. These two evidence sources are associated with different uncertainty effects and can support different responses. DUET treats them as operational evidence sources; this scope does not attempt to characterize every uncertainty mechanism in a VLA policy.

DUET formulates monitoring as two measurements with explicit evidence sources. The visual branch uses demonstration-relative novelty as an evidence signal associated with epistemic effects. The action-trace branch uses local conditional action dispersion as an evidence signal associated with aleatoric effects, given an observation and a noisy flow state. At each decision prefix, the two readouts are computed, referenced, and calibrated separately. Here, decoupled denotes this operational separation; it does not claim statistical independence or a complete characterization of policy uncertainty. The calibrated scores feed a common EXECUTE/CAUTION/HELP interface.

The visual target is computed from pre-fusion policy tokens. A density model fitted to successful demonstrations yields a scene-level novelty score, while the token scores provide corresponding image locations. The action target is the trace of a local conditional action covariance. Under Gaussian flow matching, a covariance–divergence identity expresses this target through the local geometry of the action velocity field. DUET evaluates the frozen field on the complete effective action chunk using Hutchinson probes and reverse-mode vector–Jacobian products (VJPs), then calibrates the resulting readout on successful prefixes.

The evaluation proceeds from controlled action tests to visual detection and deployment. In same-policy and matched-input route controls, the action readout attains AUROCs of 0.963 ± 0.052 and

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.

0.976 ± 0.033, respectively. Under matched input, the visual control remains at AUROC 0.500. The action ordering is preserved on new scenes and across training seeds. Held-out suites reach 0.940 ± 0.006 frame AUROC and 0.970 ± 0.003 episode AUROC, and the same token statistic provides spatial evidence. We further evaluate fixed-threshold routing and physical visual HELP alerts.

Our contributions are threefold. First, demonstration-relative visual novelty and local conditional action dispersion define distinct monitoring targets with explicit statistical scope. Second, frozen visual tokens and local action-field derivatives provide practical readouts for these targets, supported by matched-input controls, repeated protocols, and complete-action audits. Third, separate calibration preserves the source of each signal when the readouts are mapped to response policies, as shown by detection, localization, frozen routing, and physical visual alerts. This formulation keeps alert provenance available to downstream decisions.

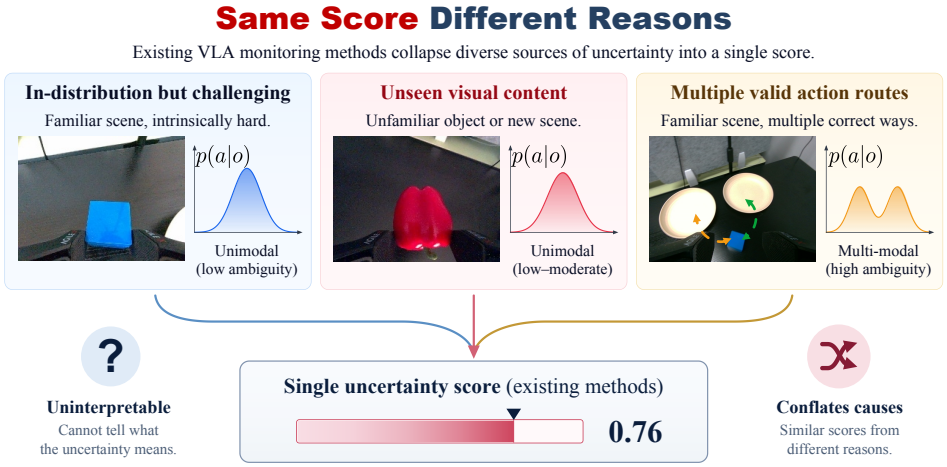

Figure 1: Source-aware cases: visual novelty for unseen content and action dispersion for multiple valid routes.

2 Related Work

Monitoring for VLA policies. Recent VLA monitors study task failure, success confidence, and action consistency. Vision-language-action policies build on multimodal robot-learning systems such as RT-2 (Zitkovich et al., 2023) and the cross-embodiment RT-X family (Open X-Embodiment Collaboration, 2024). SAFE learns failure signals across tasks (Gu et al., 2025), while DUET reads evidence directly from the frozen policy field. DUET studies the evidence available at a decision prefix of a frozen policy. It constructs a visual reference from successful demonstrations, evaluates local geometry in the action field, and calibrates the two channels independently. Comparisons in Section 4 retain each method's supervision and sampling protocol.

Uncertainty and generative action models. Monitoring systems often distinguish experience coverage from variation in generated output (Kendall & Gal, 2017; Hüllermeier & Waegeman, 2021; Ovadia et al., 2019). DUET uses this distinction as an operational guide: demonstration-relative visual novelty is associated with epistemic effects, and local action-trace variation with aleatoric effects. Evidential models estimate uncertainty through learned evidence, with semantics determined by their representation and training objective (Sensoy et al., 2018; Charpentier et al., 2022; Bengs et al., 2022; Shen et al., 2024). DUET uses the Gaussian flow-matching construction (Lipman et al., 2023) and stochastic trace estimation (Hutchinson, 1989) to construct a local action readout inside a frozen VLA. DUET evaluates this readout on the complete action chunk, fixed interpolation states, and a matched-input intervention within the deployment protocol.

Visual references and spatial evidence. Density references over learned representations are widely used for novelty and out-of-distribution detection. DUET uses a PCA projection and a full-covariance Gaussian mixture over policy tokens, with the reference fitted from successful demon-

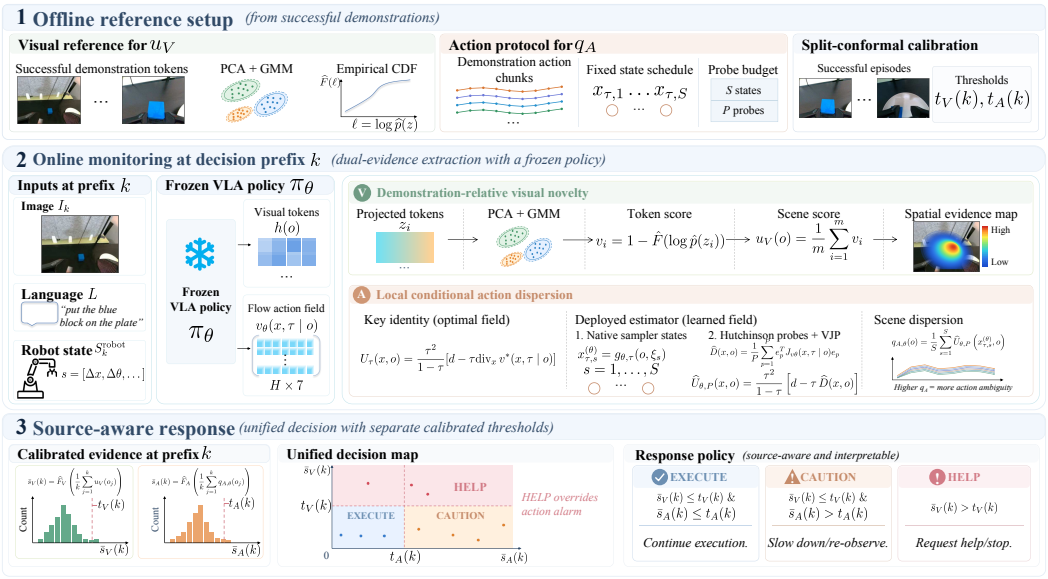

Figure 2: Overview of DUET. The visual evidence branch uses demonstration-relative novelty, while the action evidence branch uses local conditional dispersion. These readouts are associated with epistemic and aleatoric effects, respectively. Separate successful references calibrate the two readouts before they enter the response interface.

strations. Standard feature-space references provide comparison points (Bishop, 2006; Mahalanobis, 1936; Oquab et al., 2024). The same token statistic yields a scene-level score and a spatial evidence map. These outputs are evaluated in scene detection, region localization, and the physical visual-HELP protocol.

Calibration and response interfaces. Split-conformal calibration supplies operating points for fixed scores under exchangeability (Angelopoulos & Bates, 2023; Guo et al., 2017). DUET applies it separately to visual and action prefix statistics. The resulting percentiles define a configurable EXECUTE, CAUTION, and HELP interface whose priority follows the task protocol. The interface retains the two calibrated evidence streams.

3 DUET: Source-Preserved Monitoring in Frozen VLA Policies

We consider a frozen flow-based VLA. Let o denote an observation containing an image, an instruction, and robot state, and let $a \in \mathbb{R}^d$ denote the effective action vector for one action chunk, with $d = H \times 7$. Policy parameters remain fixed during deployment. Reference distributions and calibration thresholds are fitted offline from successful demonstrations. DUET exposes two source-aware evidence readouts: the visual evidence branch reports demonstration-relative novelty $u_V(o)$, and the action evidence branch reports local conditional action dispersion $q_A(o)$. We use these readouts as signals associated with epistemic and aleatoric effects, respectively. The action readout uses every effective coordinate in the complete action chunk.

At each monitored prefix, the two readouts are computed, referenced, and calibrated separately. These readouts retain two distinct evidence sources. The visual readout measures demonstration-relative novelty, and the action readout measures residual action variation conditioned on the observation and noisy flow state. Each target has an explicit reference, score, and calibration procedure. The terms epistemic and aleatoric describe the motivating uncertainty effects associated with these branch roles; they do not characterize every uncertainty mechanism in a policy.

3.1 MONITORING TARGETS AND THEIR STATISTICAL SCOPE

For the visual branch, the target is demonstration-relative novelty used as a signal associated with epistemic effects. At deployment, tokens are evaluated against a successful-demonstration reference tied to the frozen representation, PCA projection, and density model. Lower reference density yields higher novelty. The same token scores provide scene-level and spatial evidence.

For the action branch, the target is local conditional action dispersion used as a signal associated with aleatoric effects.

$$U_\tau(x, o) := \operatorname{tr} \operatorname{Cov}(a \mid x_\tau = x, o). \tag{1}$$

This quantity measures action variation after the observation and complete noisy flow state have been fixed. It quantifies residual action variation after conditioning on the current state. We use the Gaussian flow-matching construction (Lipman et al., 2023)

$$x_\tau = (1-\tau)a + \tau\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I_d), \qquad v^*(x, \tau \mid o) = \mathbb{E}[\epsilon - a \mid x_\tau = x, o], \tag{2}$$

where $0 < \tau < 1$ and the noise is independent of (a, o). For the optimal differentiable velocity field,

$$U_\tau(x, o) = \frac{\tau^2}{1-\tau} \left[d - \tau \operatorname{div}_x v^*(x, \tau \mid o) \right]. \tag{3}$$

Appendix B gives the assumptions and derivation. The identity makes the action evidence readout accessible through action-field geometry.

Predictive covariance provides a complementary distribution-level summary:

$$\operatorname{tr} \operatorname{Cov}(a \mid o) = \mathbb{E}_{x_\tau \mid o}[U_\tau(x_\tau, o)] + \operatorname{tr} \operatorname{Cov}_{x_\tau \mid o}(\mathbb{E}[a \mid x_\tau, o]). \tag{4}$$

The second term captures changes in the conditional mean across flow states. Equation (4) places the local target within the full action distribution. DUET evaluates the complete effective action vector at specified states and aggregates its local readout. Route count and local conditional dispersion therefore provide complementary descriptors. The route experiments use valid single- and two-route conditions under a fixed interpolation-state protocol to evaluate the learned-field ordering on the complete action readout.

The following proposition explains why the visual reference is defined separately.

Proposition 1 (Separated off-support contexts)**.** *Fix $\tau \in (0, 1)$. Let S_τ be the closed support of ID context–interpolant pairs $z = (c, x)$, and let v_0 be a smooth field. If $\operatorname{dist}(z_0, S_\tau) > 0$, then for every $b \in \mathbb{R}$ there exists a smooth field v_b with exactly the same ID flow loss as v_0, but whose divergence at z_0 differs from that of v_0 by b.*

The proof constructs a smooth perturbation outside the ID support (Appendix B). Gaussian interpolation has full action–state support for a supported context, so the separation is carried by the context coordinate. Thus, the visual reference remains tied to successful-demonstration coverage.

3.2 FROZEN-POLICY READOUTS

DUET uses two accessible components of the frozen policy: pre-fusion visual tokens and the action velocity field. For the visual branch, a PCA projection reduces token dimension and a full-covariance Gaussian mixture estimates the reference density $\hat{p}$. A disjoint reference set defines the empirical CDF $\hat{F}$ of log densities. For projected tokens $z_1, \ldots, z_m$,

$$\nu_i = 1 - \hat{F}\big(\log \hat{p}(z_i)\big), \qquad u_V = \frac{1}{m} \sum_{i=1}^{m} \nu_i. \tag{5}$$

Low-density tokens receive high novelty values. Their mean gives the scene-level readout, and their original grid positions give the spatial map. The same token statistic supports both scene detection and spatial evidence.

For the action branch, we apply the optimal-field functional to the trained field v_θ and estimate the divergence with Hutchinson probes:

$$\widehat{D}(x, o) = \frac{1}{P} \sum_{p=1}^{P} e_p^\top J_{v_\theta} e_p, \qquad \widehat{U}_\tau(x, o) = \frac{\tau^2}{1-\tau} \left[d - \tau \widehat{D}(x, o) \right], \tag{6}$$

where $e_p \in \mathbb{R}^d$ are independent Rademacher directions. The learned-field quantity $\widetilde{U}_\theta$ is the exact-divergence functional, while $\widehat{U}_{\theta,P}$ is the finite-P estimator used by the deployed score. Their relation to the population target is

$$\widetilde{U}_\theta - U_\tau = -\frac{\tau^3}{1-\tau}\,\mathrm{div}_x(v_\theta - v^*). \tag{7}$$

We use this correction term to interpret the numerical and operational-reference checks in Section 4. Accordingly, the deployed action score is a learned-field proxy motivated by the population identity and calibrated on successful prefixes. The main route experiments use the complete native 50×7 effective action chunk.

Reverse-mode vector–Jacobian products evaluate each quadratic form through $e_p^\top \nabla_x (e_p^\top v_\theta)$, following stochastic trace estimation (Hutchinson, 1989). We additionally compare the VJP computation with a finite-difference implementation. Both implementations reuse the visual-language prefix.

The deployed action readout averages its field statistic over fixed interpolation states. The frozen SmolVLA sampler uses $N = 10$ native Euler updates with $\Delta\tau = -1/N = -0.1$ on $\tau_j = 1 - j/N$. The state before update $j = 5$ is exactly at $\tau = .5$, and the VJP call receives this time value. A fixed seeded Gaussian noise list gives the $S = 8$ states in the primary route protocol. These native Euler states define the online readout. The Gaussian interpolation target in Eq. (2) and the fixed-state reverse-SDE reference serve their respective mathematical and audit purposes. Paired conditions reuse noise seeds, but each model integrates its own learned field, so the resulting states remain model-specific. The visual-language prefix is cached once; state construction and suffix evaluations are included in the monitored pass.

$$x_{\tau,s}^{(\theta)} = g_{\theta,\tau}(o, \xi_s), \qquad q_{A,\theta}(o) = \frac{1}{S}\sum_{s=1}^{S} \widehat{U}_{\theta,P}\left(x_{\tau,s}^{(\theta)}, o\right), \tag{8}$$

where $g_{\theta,\tau}$ denotes the native Euler sampler induced by the frozen model θ and ξ_s is its fixed noise seed. The primary action coordinates are the seven physical output features per chunk step; SmolVLA padding coordinates are masked, and all values use the frozen policy's normalized action convention. We retain raw finite-P values through aggregation and calibration. Thus, the reference, threshold, and held-out scores share one ordinal scale. The controlled route experiments use $(\tau, P, S) = (.5, 4, 8)$ with VJP. Section 4 reports complete-action route ordering, and Appendix D.1 reports the $d = 350$ operational-reference audit.

3.3 Separate calibration and response interface

The two readouts use separate successful-ID references. Let $\tilde{s}_V = u_V$ and $\tilde{s}_A = q_A$ denote raw per-prefix readouts. For channel $c \in \{V, A\}$, the interface forms the prefix statistic used at test,

$$\bar{\tilde{s}}_c(k) = k^{-1}\sum_{j=1}^{k} \tilde{s}_c(j), \qquad \bar{s}_c(k) = \widehat{F}_c(\bar{\tilde{s}}_c(k))\,,$$

where $\widehat{F}_c$ is the successful-ID reference CDF. Interface thresholds are percentile thresholds on $\bar{s}_c$, while raw q_A and u_V remain the quantities used in audits and readout comparisons. Calibration uses the same prefix length and score definition as held-out evaluation. For n exchangeable calibration units, the split-conformal threshold is the order statistic at rank

$$r = \lceil (n+1)(1-\alpha) \rceil. \tag{9}$$

We set the threshold to $+\infty$ when $r > n$ and use strict exceedance (Angelopoulos & Bates, 2023; Guo et al., 2017). The operating point is defined at the specified decision prefix. The calibration unit is the episode prefix at the same chunk index used at test time.

The response rule is

$$R_k = \begin{cases} \text{HELP}, & \bar{s}_V(k) > t_V(k), \\ \text{CAUTION}, & \bar{s}_V(k) \le t_V(k) \text{ and } \bar{s}_A(k) > t_A(k), \\ \text{EXECUTE}, & \bar{s}_V(k) \le t_V(k) \text{ and } \bar{s}_A(k) \le t_A(k). \end{cases} \tag{10}$$

Visual novelty has priority because the physical protocol uses HELP to stop execution before irreversible motion. CAUTION is the task-defined response to a high action readout under a visual non-alarm. EXECUTE indicates that both calibrated readouts remain within their operating ranges at the current prefix. The interface is configurable: response actions can be attached to these labels while the two evidence sources and their calibration references remain explicit.

4 EXPERIMENTS

We organize the experiments around the proposed evidence chain. We first evaluate the action readout with same-policy and matched-input route controls, including repetitions on new scenes and training seeds. We then evaluate visual novelty through held-out-suite detection and spatial localization. Finally, we examine numerical consistency, state averaging, deployment cost, fixed response operating points, and the physical visual-HELP protocol. Except for the offline quota-matching diagnostic in Section C.8, thresholds are selected from the stated calibration split and fixed before held-out evaluation.

4.1 EXPERIMENTAL SETTING

We use SmolVLA (Shukor et al., 2025) for LIBERO (Liu et al., 2023) and for the physical-robot study. Policy weights remain frozen during monitoring. The visual head uses 64-dimensional PCA token features and a 32-component full-covariance Gaussian mixture. Fitting, reference, calibration, and test data are disjoint. The action head uses the complete effective action vector defined in Section 3, with $(\tau, P, S) = (.5, 4, 8)$ as the primary setting.

Both branches use full-FP32 inference, fixed inputs, disjoint assets, and thresholds fixed before held-out evaluation; the separate offline quota-matching diagnostic is defined in Appendix C.8. The visual protocols evaluate the visual branch under changed visual support while retaining the policy input format. The action protocols evaluate the action-trace branch under changed demonstrated action distributions while matching the policy input. These interventions test the branch associations used by DUET. The experiments evaluate visual detection and localization, action ordering under valid routes, repeatability across scenes and seeds, and response-interface consistency. Unless stated otherwise, $\pm$ denotes standard deviation across fitted-model seeds and brackets give 95% intervals.

4.2 ACTION READOUT UNDER CONTROLLED ROUTES

The controlled testbed contains a gripper, a target, and an obstacle. A one-route condition has 16 demonstrations on one permitted side. A two-route condition has eight demonstrations on each side. Every route passes the geometric validity checks, and instructions are complete and identical within each paired comparison.

The same-policy protocol uses one frozen checkpoint. It compares pre-commitment contexts with identical image, state, and complete instruction, paired with balanced left/right successful demonstrations, against post-commitment contexts whose robot state is already on one side and whose continuation demonstrations are single-mode. The state and route cue distinguish the context records. Each seed uses 96 training, 40 calibration, and 48 held-out base scenes. The VJP trace with $P = 4$ and $S = 8$ reaches 0.963 ± 0.052 AUROC. In the matched-input control, the image, instruction, state, route cue, obstacle geometry, and demonstration count are fixed. Two models use matched initializations: one receives eight left and eight right successful demonstrations, while the other receives 16 demonstrations on one route. The vision-language backbone remains frozen, and the action expert evaluates the complete native 50×7 chunk. The trace reaches 0.976 ± 0.033 AUROC, while the visual control remains fixed at AUROC 0.500.

The matched-input protocol therefore isolates sensitivity to the fitted action distribution while holding visual input and route geometry fixed. Valid samples from the two-mode model cover both permitted routes in every held-out context.

Table 1: Action-mode discrimination on the controlled valid-route testbed.

Control	Readout	P	S	AUROC ↑
Same policy	VJP trace	4	8	$\mathbf{.963 \pm .052}$
Matched input	VJP trace	4	8	$\mathbf{.976 \pm .033}$
Matched input	Visual novelty	—	—	$.500 \pm .000$

The deployed trace is evaluated over the complete effective action vector in the route experiments. Numerical fidelity, operational-reference alignment, state averaging, and runtime are consolidated in Section 4.4; full sampler definitions and audit configurations are given in Appendix D.1.

We repeat the full-FP32 protocol with thresholds calibrated on the corresponding single-route reference scenes and frozen for each confirmation protocol. The fixed-checkpoint study uses training seeds 111–113 and a new scene seed. The P4/S8 readout has AUROC 1.000 for every checkpoint, with 96.71% mean mixture validity, a 100% mixture trigger rate, and a 1.74% single-route trigger rate. A second study trains six new checkpoints from the SmolVLA base with seeds 3101–3106. Training, calibration, and test scenes are new. Every checkpoint reaches P4/S8 AUROC 1.000, with 97.78% mean mixture validity, a 100% mixture trigger rate, and a 3.30% single-route trigger rate. These rates indicate selective triggering under both confirmation protocols.

Table 2: Frozen action readout on new scenes and new training seeds. N denotes the number of checkpoints; route thresholds are selected from single-route calibration scenes.

Protocol	N	AUROC	Single-route	Mixture	Validity
Fixed / new scenes	3	$\mathbf{1.000}$	1.74%	$\mathbf{100}\%$	$\mathbf{96.71}\%$
New seeds / new scenes	6	$\mathbf{1.000}$	3.30%	$\mathbf{100}\%$	$\mathbf{97.78}\%$

4.3 Visual novelty and spatial evidence

LOSO evaluation holds out one LIBERO suite at a time. DUET reaches 0.940 ± 0.006 frame AUROC and 0.970 ± 0.003 episode AUROC. The shared-fit GMM–Mahalanobis baseline reaches 0.916 ± 0.025 mean frame AUROC under the same tokens, folds, seeds, and feature budget.

Table 3: LOSO frame AUROC. All feature-based methods use the same demonstration-feature setting.

Method	Spatial	Object	Goal	Long	Mean
DUET	$\mathbf{.854 \pm .026}$	$\mathbf{.999 \pm .001}$	$\mathbf{.908 \pm .002}$	$\mathbf{.999 \pm .000}$	$\mathbf{.940 \pm .006}$
GMM–Maha	$.821 \pm .066$	$.998 \pm .001$	$.881 \pm .041$	$.966 \pm .002$	$.916 \pm .025$
kNN cosine	$.587 \pm .008$	$.833 \pm .029$	$.722 \pm .065$	$.894 \pm .008$	$.759 \pm .021$
kNN L_2	$.496 \pm .010$	$.918 \pm .009$	$.767 \pm .013$	$.759 \pm .005$	$.735 \pm .003$
Mahalanobis	$.244 \pm .011$	$.851 \pm .008$	$.685 \pm .004$	$.596 \pm .009$	$.594 \pm .003$
PCA reconstruction	$.252 \pm .003$	$.893 \pm .005$	$.709 \pm .013$	$.518 \pm .013$	$.593 \pm .003$

The token map also localizes changed regions. Region AUROC is $.990 \pm .002$ for synthetic changes and $.892 \pm .007$ for masked changes. The masked matched-compositing control reaches $.835 \pm .004$. On 120 wrist-camera frames, an unfamiliar red pepper receives a higher score than a familiar blue block in 90% of comparisons, with region AUROC .919. We define Localization as the top-3 token hit rate against the affected-token set; Region AUROC summarizes token-level ranking.

Table 4: Spatial localization with the visual evidence map. Localization is top-3 token hit rate against the annotated affected-token set; Region AUROC is computed over token scores.

Setting	Region AUROC ↑	Localization ↑	Matched control AUROC
Synthetic	$\mathbf{.990 \pm .002}$	$\mathbf{.944 \pm .013}$	—
Masked	$\mathbf{.892 \pm .007}$	$\mathbf{.790 \pm .022}$	$\mathbf{.835 \pm .004}$
Hardware	**.919**	**.742**	—

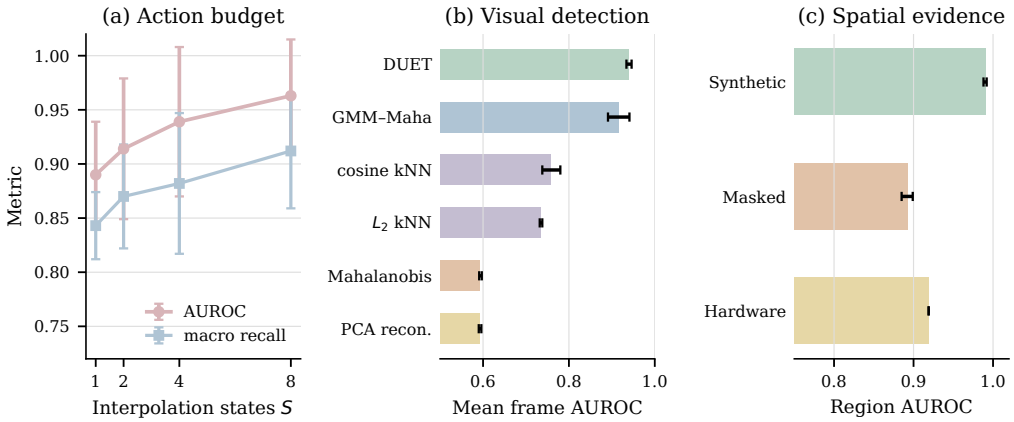

Figure 3: Readout performance and state-budget analysis. (a) Action discrimination and three-way macro recall as the interpolation-state budget increases. (b) Mean LOSO visual detection AUROC across reference baselines. (c) Region AUROC for the visual evidence map. Error bars follow the unit reported for each panel: model-seed variation for action curves, LOSO variation for visual comparison, and localization intervals for region results.

The same token statistic provides the spatial evidence reported in Table 4 and Fig. 3c. The complete feature-baseline suite is retained in Table 3.

In protocol-matched comparisons, DUET attains 0.962, compared with 0.938 for the SAFE-adapted supervised probe, and 0.939, compared with 0.730 for the evidential head under the single-restart density protocol. The comparison settings and fitting details are reported in Appendix C.5.

4.4 Numerical consistency, state budget, and deployment cost

The VJP implementation agrees with the corresponding explicit-Jacobian quadratic form. Over 768 contexts and 6144 states, the maximum absolute difference between the two quadratic-form evaluations is 2.96×10^{-7}. A finite-difference bridge on paired observations gives Spearman correlation 1.000. Mean absolute differences are 7.25×10^{-4} for one state and 4.42×10^{-4} after eight-state averaging, with a maximum absolute difference below 1.68×10^{-3}. The complete $d = 350$ audit gives Spearman .930 for one state and .940 for the primary eight-state setting, with Pearson correlations .922 and .944, respectively (Table 5).

Table 5: Full-action operational-reference audit on the effective $d = 350$ action block. The reference is the fixed-state learned-field reverse-SDE trace; P is fixed at 4.

States S	Probes P	Spearman ↑	Pearson ↑
1	4	.930	.922
8	4	.940	.944

With four VJP probes and nested subsets of the same eight-state bank, averaging additional interpolation states improves route separation and response selectivity. Across $S = 1, 2, 4, 8$, AUROC is .890, .914, .939, .963 and three-way macro recall is .843, .870, .882, .912, respectively. At $S = 8$, single-route CAUTION is 3.5% and two-route CAUTION is 86.1%. We use $S = 8$ for the primary route and routing results. Table 12 reports the corresponding warm-start deployment pass, including state construction and VJP evaluation.

The paired deployment times are 114.43 ms baseline and 150.94 ms monitored at $S = 1$, and 394.56 ms baseline and 453.76 ms monitored at $S = 8$. The matched-trajectory increments are 36.51 ms and 59.20 ms, respectively; the full timing configuration is given in Appendix D.

4.5 Fixed-threshold routing and geometric response allocation

The three response labels are assigned from fixed pooled thresholds on calibrated percentile prefix scores (0.875 for $\bar{s}_A$ and 0.975 for $\bar{s}_V$), with HELP taking priority. Source and transfer estimates are reported in Appendix C.7.

Table 6: Frozen three-way routing; pooled intervals use the data-seed/model-seed hierarchy.

Condition	All protocols	Held-out transfer protocols
Clean → EXECUTE	$\mathbf{85.14}\%$ $[\mathbf{80.52}, \mathbf{88.75}]\%$	$\mathbf{84.23}\%$
Controlled two-route mixture → CAUTION	$\mathbf{90.72}\%$ $[\mathbf{85.81}, \mathbf{95.38}]\%$	$\mathbf{88.82}\%$
OOD/mixed → HELP	$\mathbf{100}\%$	$\mathbf{100}\%$

The source-conditioned rule reaches 99.792% overall and 99.623% on the prespecified response challenge; source and transfer rows are in Appendix C.7, with the protocol in Appendix C.8.

4.6 Physical robot early visual alerts

Physical HELP fires in 45/50 unfamiliar-object trials and 1/50 familiar-object trials; all unfamiliar alerts precede irreversible motion.

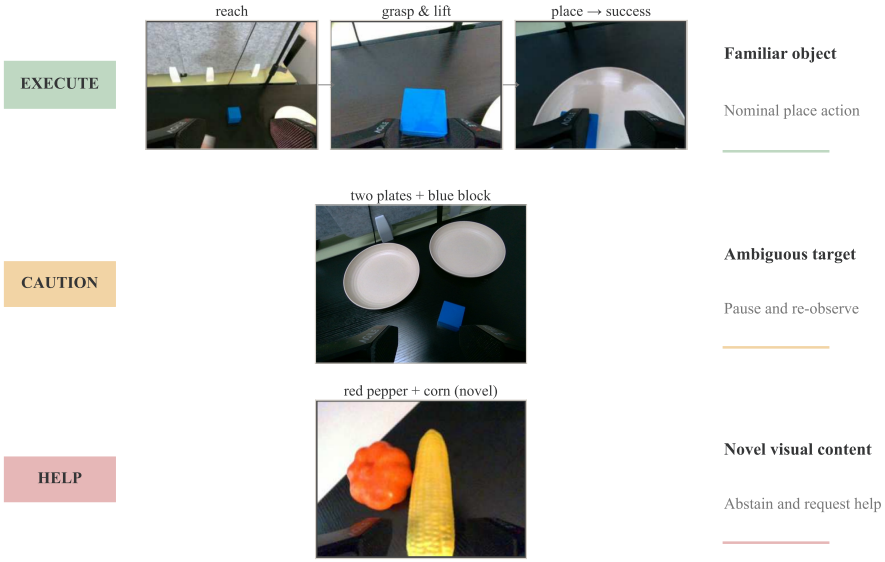

Figure 4: Physical-robot examples of the EXECUTE, CAUTION, and HELP operating points.

5 Discussion and Conclusion

DUET separates visual novelty from local action dispersion in a frozen VLA. The two readouts are complementary because they use different references and conditioning variables. Visual token novelty compares the current token distribution with a reference built from successful demonstrations, measuring whether the observed visual evidence is covered by that reference. The action trace evaluates the local geometry of the frozen action field after the observation and noisy flow state are fixed, measuring the residual variation among action continuations under the same conditioned state. These readouts therefore answer different monitoring questions and can support different responses.

The matched-input control makes this distinction explicit: the action readout changes with the fitted action distribution while the visual control remains unchanged. Separate calibration preserves this distinction when the scores enter the EXECUTE, CAUTION, and HELP interface. Controlled checks and physical HELP alerts (45/50 unfamiliar-object; 1/50 familiar-object, all before irreversible motion) support source-aware deployment.

AI USE STATEMENT

AI tools were used to adjust the manuscript layout and page formatting.

ETHICS STATEMENT

Robotic monitoring controls when a system acts and when it requests assistance. Our physical-robot protocol uses early stopping and independent response controls. Simulation experiments provide controlled visual and action interventions. The experiments evaluate the specified monitoring and response protocols.

REPRODUCIBILITY STATEMENT

Appendix B derives the statistical identity and the scope of the off-support proposition. Appendix C specifies configurations, splits, calibration, response implementations, and resampling units. Appendix D records the numerical fidelity, runtime, and reproducibility checks.

REFERENCES

Anastasios N. Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning*, 16(4):494–591, 2023. doi: 10.1561/2200000101.

Viktor Bengs, Eyke Hüllermeier, and Willem Waegeman. Pitfalls of epistemic uncertainty quantification through loss minimisation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 29205–29216, 2022.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.

Bertrand Charpentier, Oliver Borchert, Daniel Zügner, Simon Geisler, and Stephan Günnemann. Natural posterior network: Deep bayesian predictive uncertainty for exponential family distributions. In *International Conference on Learning Representations (ICLR)*, 2022.

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)*, volume 38, pp. 40041–40076, 2025. doi: 10.52202/085713-1337.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017.

Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110(3):457–506, 2021.

Michael F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 18(3):1059–1076, 1989. doi: 10.1080/03610918908812806.

Alex Kendall and Yarin Gal. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. LIBERO: Benchmarking knowledge transfer for lifelong robot learning. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2023.

Prasanta Chandra Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1):49–55, 1936.

Open X-Embodiment Collaboration. Open X-embodiment: Robotic learning datasets and RT-X models. In *2024 IEEE International Conference on Robotics and Automation*, pp. 6892–6903. IEEE, 2024. doi: 10.1109/ICRA57147.2024.10611477.

Maxime Oquab, Timothée Darcet, Thomas Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daan Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Mike Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024.

Yaniv Ovadia, Emily Fertig, Jie Ren, Zachary Nado, D. Sculley, Sebastian Nowozin, Joshua V. Dillon, Balaji Lakshminarayanan, and Jasper Snoek. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, 2018.

Maohao Shen, Jongha Jon Ryu, Soumya Ghosh, Yuheng Bu, Prasanna Sattigeri, Subhro Das, and Gregory W. Wornell. Are uncertainty quantification capabilities of evidential deep learning a mirage? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 107830–107864, 2024.

Mustafa Shukor, Dana Aubakirova, Francesco Capuano, Pepijn Kooijmans, Steven Palma, Adil Zouitine, Michel Aractingi, Caroline Pascal, Martino Russi, Andres Marafioti, Simon Alibert, Matthieu Cord, Thomas Wolf, and Remi Cadene. SmolVLA: A vision-language-action model for affordable and efficient robotics. *arXiv preprint arXiv:2506.01844*, 2025.

Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, Quan Vuong, Vincent Vanhoucke, Huong Tran, Radu Soricut, Anikait Singh, Jaspiar Singh, Pierre Sermanet, Pannag R. Sanketi, Grecia Salazar, Michael S. Ryoo, Krista Reymann, Kanishka Rao, Karl Pertsch, Igor Mordatch, Henryk Michalewski, Yao Lu, Sergey Levine, Lisa Lee, Tsang-Wei Edward Lee, Isabel Leal, Yuheng Kuang, Dmitry Kalashnikov, Ryan Julian, Nikhil J. Joshi, Alex Irpan, Brian Ichter, Jasmine Hsu, Alexander Herzog, Karol Hausman, Keerthana Gopalakrishnan, Chuyuan Fu, Pete Florence, Chelsea Finn, Kumar Avinava Dubey, Danny Driess, Tianli Ding, Krzysztof Marcin Choromanski, Xi Chen, Yevgen Chebotar, Justice Carbajal, Noah Brown, Anthony Brohan, Montserrat Gonzalez Arenas, and Kehang Han. RT-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*, volume 229, pp. 2165–2183. PMLR, 2023.

CONTENTS

1 Introduction **1**

2 Related Work **2**

3 DUET: Source-Preserved Monitoring in Frozen VLA Policies **3**
3.1 Monitoring targets and their statistical scope . . . 4
3.2 Frozen-policy readouts . . . 4
3.3 Separate calibration and response interface . . . 5

4 Experiments **6**
4.1 Experimental setting . . . 6
4.2 Action readout under controlled routes . . . 6
4.3 Visual novelty and spatial evidence . . . 7
4.4 Numerical consistency, state budget, and deployment cost . . . 8
4.5 Fixed-threshold routing and geometric response allocation . . . 9
4.6 Physical robot early visual alerts . . . 9

5 Discussion and Conclusion **9**

A Appendix Guide **13**

B Statistical Derivations and Scope **13**
B.1 Gaussian conditional-covariance identity . . . 13
B.2 Separated context proposition . . . 13
B.3 Probe estimator . . . 13

C Protocols and Statistical Units **14**
C.1 Visual detection and localization . . . 14
C.2 Controlled routes and state averaging . . . 14
C.3 Matched-input details . . . 14
C.4 Interpolation-state budget . . . 15
C.5 Visual head comparisons and transfer . . . 15
C.6 New scenes and new training seeds . . . 15
C.7 Three-way routing . . . 15
C.8 Geometric response allocation . . . 16

D Implementation, Fidelity, and Reproducibility **16**
D.1 Full-action target and operational-reference construction . . . 16
D.2 Implementation and numerical reproducibility . . . 18

E Physical-Robot Protocol **18**

A Appendix Guide

The appendix is organized by the evidence chain established in the paper. Appendix B fixes the statistical target and derives the flow-field identity that motivates the action readout. Appendix C records the data partitions, evaluation units, calibration rules, and controlled route protocols. Appendix D reports numerical fidelity, runtime, and reproducibility checks. Appendix E gives the physical-robot protocol and the per-trial deployment details.

B Statistical Derivations and Scope

The appendix evaluates the visual-novelty and local-conditional-dispersion targets under the calibration protocols and evaluation units defined below. All reported statistics are operational readouts and are interpreted within their stated protocols.

B.1 Gaussian conditional-covariance identity

Fix o and write $\beta = 1 - \tau$. Assume $a \mid o$ has finite second moments and differentiation can pass through the Gaussian convolution. For $0 < \tau < 1$, let

$$x_\tau = (1-\tau)a + \tau\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I_d). \tag{11}$$

Writing $m(x) = \mathbb{E}[a \mid x_\tau = x, o]$, differentiation gives

$$J_m(x) = \frac{\beta}{\tau^2}\,\mathrm{Cov}(a \mid x_\tau = x, o). \tag{12}$$

For the optimal flow field $v^*(x, \tau \mid o) = \mathbb{E}[\epsilon - a \mid x_\tau = x, o]$,

$$\mathrm{Cov}(a \mid x_\tau = x, o) = \frac{\tau^2}{1-\tau}\left(I_d - \tau J_{v^*}(x, \tau \mid o)\right). \tag{13}$$

Taking the trace gives Eq. (3). The identity is exact for the optimal field; the deployed readout replaces v^* with the frozen learned field and estimates its divergence with probes. The result links local action-field geometry to conditional action dispersion.

The law of total covariance gives Eq. (4) for the complete effective action vector. DUET uses the conditional term as its action target. The implementation reads the learned field with VJP probes and calibrates the resulting score against successful reference episodes.

B.2 Separated context proposition

Fix $\tau \in (0, 1)$. Let S_τ be the closed support of ID context–interpolant pairs $z = (c, x)$. If $\mathrm{dist}(z_0, S_\tau) > 0$, then for every $b \in \mathbb{R}$ there exists a smooth field v_b with the same ID flow loss as v_0 and a divergence at z_0 shifted by b.

Choose a smooth bump χ supported in a ball around z_0 and equal to one on a neighborhood of z_0, and define

$$g_b(c, x) = \frac{b}{d}\chi(c, x)(x - x_0), \qquad v_b = v_0 + g_b. \tag{14}$$

The perturbation vanishes on the ID support and has Jacobian $(b/d)I$ at z_0. This establishes the stated divergence shift. The result motivates a separate demonstration-relative visual reference model.

B.3 Probe estimator

For an exact Jacobian and independent Rademacher directions, $\mathbb{E}[ee^\top] = I$ gives

$$\mathbb{E}[e^\top J e] = \mathrm{tr}(J). \tag{15}$$

DUET averages P probe estimates over S interpolation states, then applies the reference percentile. The route experiments use VJP with $P = 4$ and $S = 8$.

C PROTOCOLS AND STATISTICAL UNITS

C.1 VISUAL DETECTION AND LOCALIZATION

The visual head extracts pre-fusion policy tokens, projects them to 64 dimensions with PCA, and fits a 32-component full-covariance GMM. The PCA and density parameters use successful demonstration tokens, while a disjoint 20% token reference split defines the empirical density percentile. Each LOSO fold evaluates 509 episodes, including 130 episodes from the held-out suite, under a matched 1.5×10^5-token budget for all feature baselines. Episode AUROC averages token percentiles within an episode. Calibration seeds are separate from the three evaluation seeds. The visual score reaches 0.940 ± 0.006 frame AUROC and 0.970 ± 0.003 episode AUROC. Region AUROC reaches $.990 \pm .002$ for synthetic changes, $.892 \pm .007$ for masked changes, and .919 on the wrist-camera set. These values use frame, episode, and region units respectively, matching the evaluation granularity of each visual task.

Protocol assets and separation units for the visual study.

Asset	Unit / size	Separation rule
Policy checkpoint	one frozen SmolVLA checkpoint	not refit during monitoring
PCA/GMM fit	successful-demonstration tokens	held-out suite excluded; feature baselines use matched 1.5×10^5 token budget
CDF reference	token-level 20% split	tokens disjoint from fit pool; source episodes may overlap because split is token-level
Threshold calibration	episode prefixes at fixed chunk index	separate from evaluation episodes; score and threshold frozen before test
Evaluation	509 episodes/fold (130 held-out-suite)	held-out episodes used only for evaluation; frame/episode/region units stated above

For each LOSO fold, the held-out suite is excluded from the successful-demonstration reference. The protocol-asset table lists the unit and separation rule for each visual-study asset. The policy checkpoint, PCA/GMM fit, token-CDF reference, calibration prefixes, and evaluation episodes are treated as separate protocol units. The 20% CDF split is token-level within the reference pool; calibration prefixes and evaluation episodes are held out under the stated protocol. Visual novelty is therefore a demonstration-relative coverage score for the specified reference, and the LOSO result measures suite-shift detection under that reference.

C.2 CONTROLLED ROUTES AND STATE AVERAGING

The complete instruction is fixed across paired conditions. A single-route condition contains 16 demonstrations on one permitted side. A two-route condition contains eight demonstrations on each side. Image, text, state, route cue, and demonstration count remain matched in the matched-input control.

The main setting is $\tau = .5$, VJP, $P = 4$, and $S = 8$. For each observation, the native SmolVLA sampler uses $N = 10$ Euler updates, $\Delta\tau = -0.1$, and grid points $\tau_j = 1 - j/10$; the pre-update state at $j = 5$ is exactly $\tau = .5$, and the derivative call uses $\tau = .5$. A fixed list of Gaussian noise seeds supplies the S trajectories. Paired conditions reuse the same seeds, while each model generates its own learned-field trajectory, so this is not a shared cross-model state bank. Prefix computation is cached before suffix VJP calls. The same-policy AUROC is $.963 \pm .052$. The matched-input AUROC is $.976 \pm .033$. The finite-difference check uses $P = 2$, $S = 1$, and central-difference step 10^{-3}; its timing and numerical role are reported separately from the primary VJP route setting.

C.3 MATCHED-INPUT DETAILS

The matched-input study fits two action distributions from matched initializations while keeping the visual-language input fixed. One distribution receives eight left and eight right demonstrations. The other receives 16 demonstrations on one route. The visual-language backbone remains frozen. The two-mode condition covers both permitted route modes in every held-out context.

Table 7: Matched-input action-distribution readout and visual control.

Readout	Two modes	One mode	AUROC
VJP trace, $P = 4, S = 8$	$\mathbf{9.199 \pm 2.573}$	$\mathbf{4.263 \pm .131}$	$\mathbf{.976 \pm .033}$
Visual novelty	.48778	.48778	.500

The unchanged visual novelty shows that the intervention preserves the policy input while changing the fitted action distribution. The VJP trace quantifies local action-field dispersion at the probed observation–state pair. The result supports sensitivity to the fitted action distribution while retaining a separate visual channel.

C.4 Interpolation-state budget

The action readout improves as it averages more interpolation states. The table reports the same VJP probe bank with $P = 4$ across the controlled route protocol.

Table 8: Action readout as a function of interpolation-state budget.

States	AUROC	Single CAUTION	Multi CAUTION	Three-way macro recall
1	$.890 \pm .049$	$.118 \pm .010$	$.736 \pm .086$	$.843 \pm .031$
2	$.914 \pm .065$	$.090 \pm .010$	$.792 \pm .136$	$.870 \pm .048$
4	$.939 \pm .069$	$.069 \pm .026$	$.806 \pm .174$	$.882 \pm .065$
8	$\mathbf{.963 \pm .052}$	$\mathbf{.035 \pm .026}$	$\mathbf{.861 \pm .139}$	$\mathbf{.912 \pm .053}$

The eight-state readout gives the strongest action separation and the highest three-way macro recall in this budget sweep. Single-route CAUTION falls from 11.8% to 3.5%, while two-route CAUTION rises from 73.6% to 86.1%. The same setting is used for the primary route and routing results.

C.5 Visual head comparisons and transfer

The policy-token density head is compared with standard feature-space references under the same LOSO folds, seeds, and token budget. DUET reaches mean frame AUROC $.940 \pm .006$, compared with $.916 \pm .025$ for GMM–Mahalanobis, $.759 \pm .021$ for cosine kNN, $.735 \pm .003$ for L_2 kNN, $.594 \pm .003$ for Mahalanobis, and $.593 \pm .003$ for PCA reconstruction. The shared-fit GMM–Mahalanobis baseline reuses the same PCA projection and fitted mixture components and scores component-wise covariance-normalized distance; all baselines use the fixed orientation in which larger values indicate greater novelty. A separate transfer comparison with a SAFE-adapted supervised probe gives .962 for DUET and .938 for the adapted probe; this is a protocol-matched reference comparison rather than a reproduction claim about the original SAFE system. A single-restart density comparison gives .939 for DUET and .730 for an evidential head; Both secondary comparisons use the protocol-matched visual reference.

C.6 New scenes and new training seeds

Three fixed full-FP32 checkpoints evaluated on new scenes reach P4/S8 AUROC 1.000, with a 1.74% single-route trigger rate, a 100% mixture trigger rate, and 96.71% mean mixture validity. Six new training seeds evaluated on new scenes also reach P4/S8 AUROC 1.000, with a 3.30% single-route trigger rate, a 100% mixture trigger rate, and 97.78% mean mixture validity. For each protocol, the route threshold is selected from its single-route calibration scenes and then frozen on the new-scene test set.

C.7 Three-way routing

Thresholds are selected on a pooled calibration protocol and transferred unchanged to two held-out transfer protocols. The calibration protocol has separate validation and test partitions; the held-out protocols are evaluated without retuning. The calibrated percentile thresholds are 0.875 for the action prefix score $\bar{s}_A$ and 0.975 for the visual prefix score $\bar{s}_V$; they are not raw q_A or u_V values.

The running prefix statistic uses the same chunk index during calibration and test, and HELP has priority over CAUTION. The resulting pooled operating point is clean → EXECUTE at 85.14%, controlled two-route mixture → CAUTION at 90.72%, and OOD/mixed → HELP at 100%.

Table 9: Frozen three-way routing by transferred data protocol.

Protocol	Clean → EXECUTE	Mixture → CAUTION	OOD/mixed → HELP
Calibration protocol	86.96%	94.53%	100%
Transfer protocol A	80.71%	87.11%	100%
Transfer protocol B	87.74%	90.53%	100%
Held-out transfer (pooled)	**84.23%**	**88.82%**	**100%**
All protocols	**85.14%**	**90.72%**	**100%**

The pooled transfer row reports the fixed-threshold point estimate. The final row combines the calibration and held-out transfer protocols for the summary operating point reported in the main text.

C.8 Geometric response allocation

Eight independent model seeds use 96 training, 64 calibration, 256 validation, and 512 test base scenes. For each test scene, we evaluate single-route, multi-route, visual-novelty, and mixed potential outcomes. The source-conditioned rule uses $(\tau, P, S) = (.5, 4, 16)$, screens the first four generated candidates for CAUTION, and uses the prescribed geometrically checked recovery path for HELP. Its fixed order is HELP when $p_{\text{epis}} > 0.850$, otherwise CAUTION when $p_{\text{alea}} > 0.875$, otherwise EXECUTE. The response order and candidate resources are frozen on validation and then evaluated on test.

The response challenge retains a base scene only when its four precomputed potential outcomes satisfy the following conditions. Single-route EXECUTE succeeds. Multi-route EXECUTE fails and CAUTION succeeds. OOD EXECUTE and CAUTION fail and HELP succeeds. Mixed EXECUTE and CAUTION fail and HELP succeeds. This yields 466 four-condition scene blocks (1,864 potential-outcome entries). EXECUTE uses the first of eight policy trajectories; CAUTION screens the first four and executes the first checker-valid candidate; HELP uses the checker-valid expanded-clearance recovery path. The checker, goal, workspace, obstacle geometry, and candidate-generation pool are shared. This is an offline, quota-matched allocation diagnostic. The scoring family and response order are fixed on validation. Test-batch score ranks are used only to assign matched response quotas. On test, the source-conditioned rule reaches 99.792% over all test conditions and 99.623% on the response challenge. Differences use paired model-seed and scene resampling with 20,000 bootstrap draws.

Table 10: Controlled geometric response allocation with matched resources.

Population	Source-conditioned
All test conditions	**99.792%**
Response challenge	**99.623%**

D Implementation, Fidelity, and Reproducibility

The following checks relate the reported readouts to the numerical operations used at deployment. They use frozen checkpoints and the stated preprocessing and calibration conventions. State construction, probe/trace implementation, and audit-specific configurations are specified separately below.

D.1 Full-action target and operational-reference construction

The action audit reports three evidence layers. An FP32 audit over 144 archived observations uses $(\tau, P, S) = (.5, 4, 1)$ and gives Spearman correlation .901 with the matched local conditional reference, with hierarchical interval $[.825, .946]$. This fixture uses the supported four-dimensional first-

step controls and verifies ranking on that reference. The primary route study instead applies the same learned-field readout to the complete native $H \times 7$ action chunk and evaluates end-to-end route ordering under the fixed P4/S8 protocol. Thus, the four-dimensional result is scoped to its fixture, while the complete-action result is evaluated directly on all $d = 350$ effective coordinates.

For the full-action audit, $p_\theta(a \mid o)$ denotes the action distribution induced by the frozen policy's native Euler sampler from independent Gaussian base noise. At a held-out observation o, we first draw $a_0 \sim p_\theta(a \mid o)$ and construct one fixed intermediate state

$$x_\tau = (1-\tau)a_0 + \tau\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I_d), \tag{16}$$

with $\tau = .5$. Equation (17) defines the mathematical target posterior under this native policy distribution and the Gaussian interpolation kernel. The audit constructs an operational learned-field reverse-SDE reference, denoted $\pi^{\mathrm{rsde}}_{\theta,\tau}(\cdot \mid x, o)$, using the same frozen field that is queried by the deployed readout:

$$p_{\theta,\tau}(a \mid x, o) \propto p_\theta(a \mid o) \exp\left[-\frac{\|x - (1-\tau)a\|^2}{2\tau^2}\right]. \tag{17}$$

For decreasing τ, the continuous reverse-time SDE is

$$\begin{aligned} \mathrm{d}x &= \left[v_\theta(x, \tau \mid o) - \frac{\eta_\star(\tau)^2}{2}\, s_\theta(x, \tau \mid o)\right] \mathrm{d}\tau + \eta_\star(\tau)\, \mathrm{d}\overleftarrow{W}_\tau, \\ s_\theta &= -\frac{x + (1-\tau)v_\theta}{\tau}, \qquad \eta_\star(\tau)^2 = \frac{2\tau}{1-\tau}. \end{aligned} \tag{18}$$

The score expression follows the same flow-matching convention as Eq. (2); $\eta_\star$ is the time-reversal noise level used by this operational construction. In the actual implementation, with $\Delta\tau_j < 0$, we use the Euler–Maruyama update

$$x_{j+1} = x_j + b_\theta(x_j, \tau_j \mid o)\Delta\tau_j + \eta_\star(\tau_j)\sqrt{-\Delta\tau_j}\, z_j, \quad b_\theta = v_\theta - \tfrac{1}{2}\eta_\star^2 s_\theta, \quad z_j \sim \mathcal{N}(0, I_d). \tag{19}$$

Each of the $K = 64$ trajectories starts from the same fixed x_τ and uses an independent Brownian-noise stream generated from seed 0. We use 64 Euler–Maruyama steps on the $\sigma = \tau/(1-\tau)$ schedule with $\rho = 7$, and take the terminal $\tau = 0$ update deterministically. The resulting endpoints define $\pi^{\mathrm{rsde}}_{\theta,\tau}$, which serves as the operational learned-field reference for the complete-action audit.

For effective-coordinate mask $\mathcal{I}$ with $|\mathcal{I}| = 350$, the per-state reference trace is computed from the sampled endpoints $a^{(1)}, \ldots, a^{(K)}$ as

$$\widehat{U}_{\mathrm{ref}}(x, o) = \sum_{j \in \mathcal{I}} \frac{1}{K-1} \sum_{r=1}^{K} \left(a_j^{(r)} - \bar{a}_j\right)^2, \qquad \bar{a}_j = \frac{1}{K}\sum_{r=1}^{K} a_j^{(r)}. \tag{20}$$

When multiple interpolation states are audited, the states are kept separate and the traces are averaged in the same order as the deployed readout:

$$\widehat{U}^{(S)}_{\mathrm{ref}}(o) = \frac{1}{S}\sum_{s=1}^{S} \widehat{U}_{\mathrm{ref}}(x_{\tau,s}, o). \tag{21}$$

We never pool endpoints from different $x_{\tau,s}$ values, so variation in conditional means across states is not folded into the conditional trace. The 32 held-out observations are selected once with seed 0 from the held-out robot dataset; the reported intervals use 2,000 bootstrap replicates over observation units. They quantify variation over this audit set, while the 64 reverse-SDE draws quantify Monte Carlo variation conditional on each fixed state.

To directly audit the complete action dimensionality, we evaluate the reference and learned-field readout on the same 32 observations with the effective $d = 350$ coordinates. For each observation, we sample a_0 with the native $N = 10$ Euler policy sampler. We then form the audit state using Eq. (16). This endpoint-plus-noise construction differs from the native intermediate states $g_{\theta,\tau}(o, \xi_s)$ used in the route deployment protocol. We use $\tau = .5$ and central-difference step $\varepsilon_{\mathrm{fd}} = .001$. With $P = 4$, operational-reference alignment remains stable as S varies (Table 11). The one-state audit gives Spearman .930 (interval $[.816, .968]$). Four-state averaging gives .941 (interval $[.825, .978]$). The primary eight-state setting gives .940 (interval $[.821, .979]$). A closed-form Gaussian unit test

with the analytic Gaussian flow field, $a \sim \mathcal{N}(0, I_{350})$, fixed $x_{.5} = 0$, and exact target trace 175 gives 179.11 from 4096 reverse-SDE draws. This controlled test checks the sampler and the effective-coordinate convention.

Table 11: Full-action alignment on $d = 350$ effective coordinates at $\tau = .5$ using the central-difference realization with $\varepsilon_{\mathrm{fd}} = .001$. The reference is the fixed-state learned-field reverse-SDE trace defined in Eqs. (18)–(20).

States S	Probes P	Spearman ↑	Pearson ↑
1	4	.930	.922
2	4	.936	.928
4	4	.941	.943
8	4	.940	.944

The primary route readout uses the complete native 50×7 effective action chunk and $(\tau, P, S) = (.5, 4, 8)$. A separate numerical audit over 768 contexts and 6144 states gives maximum probe absolute error 2.96×10^{-7} when the implemented VJP quadratic form is compared with the corresponding explicit-Jacobian quadratic form on the same learned field. An additional paired audit uses eight observations, the same fixed states, and the same probe bank. P4 VJP and central-difference traces have Spearman correlation 1.000. Mean absolute differences are 7.25×10^{-4} for one state and 4.42×10^{-4} after eight-state averaging. The maximum absolute difference is below 1.68×10^{-3}. This provides a numerical bridge between the primary VJP implementation and the finite-difference realization used in the complete-action reference audit. The four-dimensional Spearman result is the finite-reference alignment check for that fixture, while Table 11 is the complete-action operational-reference audit.

Table 12: Warm deployment timing on RTX3090. Generation and monitoring include observation batching, cached visual/language features, full Euler state construction, and CUDA synchronization. The baseline is matched to the same S trajectory count.

Setting	Device / precision	S, P	T_{base}	T_{monitor}	$\Delta T_{\mathrm{matched}}$
Deployment pass	RTX3090 / FP32	1, 4	114.43 ms	150.94 ms	36.51 ms
Deployment pass	RTX3090 / FP32	8, 4	394.56 ms	453.76 ms	59.20 ms

All rows use the frozen blue-block checkpoint, the same 50×7 action chunk, effective $d = 350$ mask, FP32 arithmetic, cached prefix, and CUDA synchronization. The matched increment is $\Delta T_{\mathrm{matched}}(S) = T_{\mathrm{monitor}}(S) - T_{\mathrm{base}}(S)$. For $S = 8$, it is 59.20 ms (P95 61.39 ms). Generated trajectories can be reused by downstream sample-based statistics.

D.2 Implementation and numerical reproducibility

The action chunk has $H = 50$ steps and seven effective coordinates. The four-dimensional and full-action audits use separate references and are analyzed independently. Route states use the native 10-step Euler grid, with the pre-update state at $j = 5$ exactly at $\tau = .5$; the full-action audit instead forms a fixed state from a native Euler endpoint before its reverse-SDE reference. The deployed readout uses the complete vector and fixed $(\tau, P, S) = (.5, 4, 8)$ before testing.

E Physical-Robot Protocol

The hardware set contains a familiar blue block and five unfamiliar objects; the 120-frame localization study gives .919 region AUROC. The stopping protocol uses 22 successful in-distribution rollouts, $\alpha_V = .05$, $\alpha_A = .10$, and a fixed 20-chunk window, independently of simulated routing. HELP fires in 45/50 unfamiliar-object trials (Wilson interval $[78.6, 95.7]\%$) and in 1/50 familiar-object trials (interval $[0.4, 10.5]\%$). All 45 unfamiliar alerts precede irreversible motion; the median alert occurs at chunk 2.

Table 13: Physical-robot HELP protocol. Wilson intervals are unconditional; pre-action counts are among alerts.

Condition	Trials	HELP	Wilson interval	Pre-action
Familiar object	50	$1/50$	$[0.4, 10.5]\%$	—
Unfamiliar objects	50	$\mathbf{45/50}$	$[\mathbf{78.6}, \mathbf{95.7}]\%$	$\mathbf{45/45}$