PAIRED RESIDUAL ADAPTATION FOR VISION-LANGUAGE-ACTION POLICIES UNDER TRANSIENT VISUAL MISLOCALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A brief error in a target's apparent position can redirect a robot's actions and cause a placement to fail, even after the visual error disappears. We study how to reduce these failures through action-level adaptation of a frozen vision-language-action (VLA) policy. Our approach learns a lightweight residual action policy from paired predictions of the base model. Clean and visually perturbed observations of the same physical state are evaluated with matched sampling noise, and their action difference provides the correction target. Balanced zero-residual supervision discourages unnecessary changes to valid behavior. At deployment, the residual policy uses current images, robot state, and the proposed action chunk to adjust translational actions, without a clean reference or additional base-model queries. We evaluate the method with a frozen $\pi_{0.5}$ policy on LIBERO-Spatial, using brief rendering-only target displacements during placement. A fixed residual checkpoint improves task success on held-out initial states. Matched-budget comparisons favor retaining the base action as the starting point for learning; the separate closed-loop benefits of pairing and balanced weighting remain unresolved. These results support targeted action adaptation for transient visual errors, although some clean-condition performance is lost and broader generalization remains to be established.

1 INTRODUCTION

Consider a robot placing a part into a tray at a workstation. A brief shift in the tray's apparent position may direct the part away from the physical target. Restoring the image does not undo actions already executed: the part may already be off-center or too close to contact for recovery. Figure 1 illustrates this failure in a controlled bowl-placement episode. Although the rendered plate returns to its original position, the perturbed rollout fails, whereas the clean rollout from the same initial state succeeds.

Vision-language-action (VLA) policies map observations and language instructions to robot actions (Kim et al., 2025b; Black et al., 2025a; Physical Intelligence et al., 2025), with performance depending on viewpoints, initial states, and observation quality (Fei et al., 2025). We focus on transient visual mislocalization: the target's apparent position shifts while its physical position remains unchanged at the intervention. Controlled rendering allows us to vary both the magnitude and duration of this discrepancy.

PRA keeps the pretrained VLA policy frozen and learns a small residual policy that adjusts the translational components of its action proposal. The correction head uses current images, robot state, and the proposed action chunk, and runs at the base policy's native query rate.

Supervision comes from the frozen policy itself. During training, we query the base model on clean and perturbed views of the same physical state with matched sampling noise. The difference between the two translational predictions supplies the correction target, without newly collected expert action labels. Zero-residual examples supervise preservation of the base action. Clean reference observations are used only during training.

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.

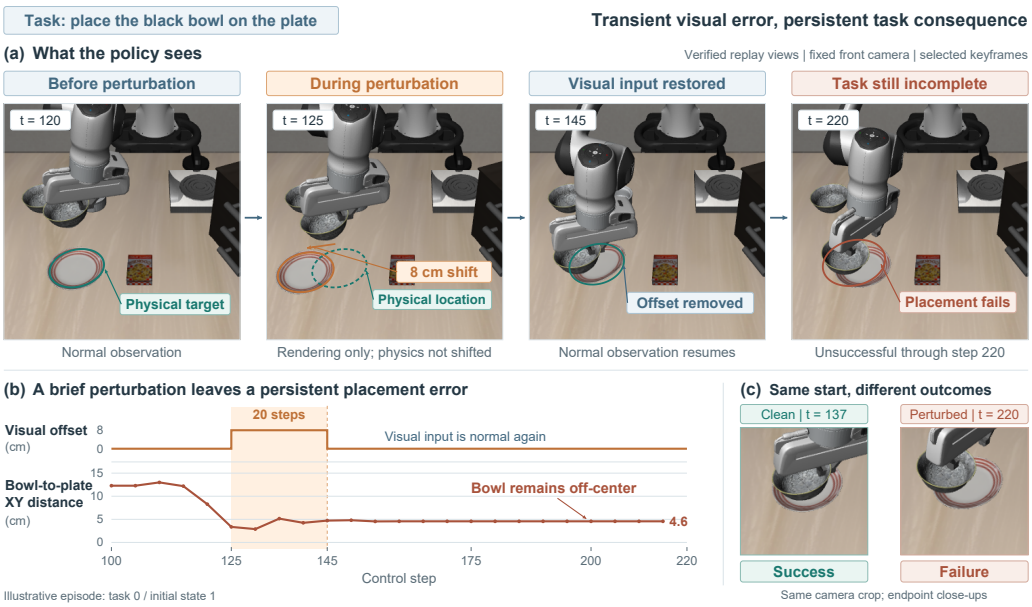


Figure 1: **A transient visual error can have a persistent task consequence.** In this illustrative development episode, the rendered plate is shifted by 8 cm at steps 125–144 and restored at step 145. The perturbed rollout remains unsuccessful through step 220; its paired clean rollout succeeds after 137 actions. Curves use logged geometric quantities, and keyframes are obtained from verified action replays at selected times.

This setting leads to three contributions. We isolate a temporary rendering error that moves the apparent target without moving its physical location at intervention, and show in a controlled rollout why restoring the image may not restore the task. We use same-state clean and perturbed predictions from a frozen VLA to train an XYZ residual policy, with zero-residual examples that encourage retention of the base action. Finally, fresh-start tests with three frozen backbones show gains under 40-step visual shifts. Matched-budget controls support retaining the base proposal as the learning starting point, and a widening 20-to-40-step advantage is confirmed for $\pi_{0.5}$.

2 Related Work

Where adaptation enters a VLA policy. OpenVLA predicts action tokens; π_0 and $\pi_{0.5}$ generate actions with a flow model (Kim et al., 2025b; Black et al., 2025a; Physical Intelligence et al., 2025). Related VLA work spans fine-tuning, test-time decoding, action tokenization, lightweight policy modules, and internal activation steering (Kim et al., 2025a; Liu et al., 2025; Pertsch et al., 2025; Wang et al., 2025; Häon et al., 2025). PRA instead corrects the completed action proposal while leaving the backbone frozen.

Correcting a proposed action. Residual policies correct an existing controller (Silver et al., 2018; Johannink et al., 2018). ResiP learns corrections through reinforcement learning; A2C2 uses demonstrations and recent observations (Ankile et al., 2024; Sendai et al., 2025). Related interventions revise contact poses from interaction feedback or verify multiple VLA proposals at test time (Xiong et al., 2025; Kwok et al., 2025). PRA instead learns an XYZ residual from paired frozen-policy predictions and applies it at the base-query rate, without online reward or candidate verification.

When visual feedback reaches control. Action chunks reduce policy queries but delay feedback. ACT and Diffusion Policy predict and execute chunks; later work adjusts their overlap or duration (Zhao et al., 2023; Chi et al., 2023; Black et al., 2025b; So et al., 2025). Runtime monitoring can flag inconsistent actions or stalled progress (Agia et al., 2025). We keep the query schedule fixed and correct the proposal when the visible target briefly shifts.

Visual perturbations and paired targets. LIBERO-Plus studies controlled perturbations on LIBERO tasks (Liu et al., 2023; Fei et al., 2025). RobustVLA, RoVLA, and Cross-View Action Consistency train policies toward stable actions across changed inputs or views (Guo et al., 2026; Luo et al., 2026; Huang et al., 2026). We instead hold the physical state fixed, displace the rendered target, and train a separate head on the frozen policy's paired action difference.

Limits of a correction target. DAgger labels learner-visited states, and DART perturbs demonstrations to expose recovery states (Ross et al., 2011; Laskey et al., 2017). Yet a better-informed teacher may still give a partially observed learner unsuitable targets (Warrington et al., 2021). Thus, our clean-policy prediction is a paired training reference, not a guaranteed recovery action.

3 Problem Formulation

3.1 Frozen Policy and Visual Intervention

Let z_t denote the physical state, o_t the front and wrist observations, and ℓ the language instruction. At a query time, the frozen policy proposes an action chunk

$$A_t = \pi(o_t, \ell; \xi_t) \in \mathbb{R}^{H \times d}, \tag{1}$$

where ξ_t is the sampling noise. Proprioceptive inputs, when used by a backbone, are implicit in π. The robot executes the first $h \leq H$ actions before querying the policy again. The added policy changes only XYZ components; the base model and its sampling procedure remain fixed. For the same physical state, the renderer provides

$$o_t^c = R(z_t; 0), \qquad o_t^v = R(z_t; \eta_t). \tag{2}$$

The intervention η_t temporarily shifts the rendered target near placement. A separate simulator produces the perturbed view without changing the physical state of the active simulator at the intervention. When clean rendering resumes, execution continues from the state reached by the actions already taken, rather than resetting to the clean trajectory.

3.2 Objective and Information Constraints

We seek fewer placement failures under visual mislocalization without needless changes to valid actions. Training uses action-reference error; evaluation uses closed-loop success. At deployment, the head receives current front and wrist images, eight-dimensional robot state s_t, and base proposal A_t, but no simulator state, clean image, target geometry, event timing, or history. Language enters through A_t. The principal learned controls share these inputs.

The head's observation interface differs from some native backbones: it adds continuous state for $\pi_{0.5}$ and wrist images and state for OpenVLA. The latter comparison therefore measures the full extension (Appendix P).

4 Paired Residual Adaptation

ResiP and Policy Decorator refine frozen policies with residuals learned from online reward (Ankile et al., 2024; Yuan et al., 2025). DAgger and DART instead collect expert-labeled or perturbed demonstrations (Ross et al., 2011; Laskey et al., 2017). Neither directly supplies a target for a brief rendering-only error in a frozen VLA. Here the same physical state yields clean and shifted views: their matched-noise XYZ action difference trains the residual head, while zero-target records ask it to preserve other base proposals (Figure 2). The clean action is a policy reference, not an expert recovery label.

4.1 Same-State Paired Action Targets

For each paired state, we query π on o^c and o^v with the same instruction and noise ξ. Let P select translation and N be fixed coordinate-wise normalization. The altered-view target is

$$\delta^v = P[N(\pi(o^c, \ell; \xi)) - N(\pi(o^v, \ell; \xi))]\,. \tag{3}$$

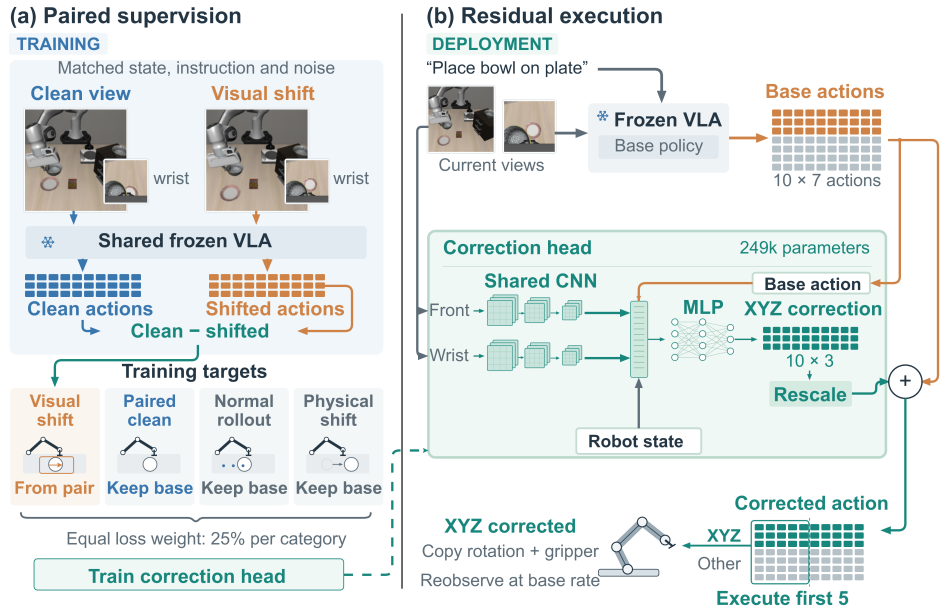


Figure 2: **Paired Residual Adaptation.** Clean and shifted views of one physical state, queried with matched noise, define an XYZ correction target. Paired clean, ordinary-rollout, and physical-shift records have zero-residual targets; the four categories are equally weighted. At deployment, the head adjusts XYZ using current images, robot state, and the proposed action chunk. Other action coordinates pass through unchanged. Only the head is trained; clean references and event timing are not deployment inputs. Camera examples are from development data, icons are schematic, and dimensions are for $\pi_{0.5}$.

We use $N(a)_k = 2(a_k - q_{0.01,k})/(q_{0.99,k} - q_{0.01,k} + \epsilon) - 1$, with $\epsilon = 10^{-6}$ and fixed archived action quantiles. Head inputs and learned controls use the same normalization. Sharing ξ avoids mixing view changes with independent sampling variation. The target describes the base policy's action change when an altered view is replaced by a clean one.

Writing the normalized base translation as $b = PN(\pi(o^v, \ell; \xi))$ and the clean reference as $y = PN(\pi(o^c, \ell; \xi))$ gives $\delta^v = y - b$. Under squared loss, regressing $f_\theta(x)$ to $y - b$ is equivalent to regressing $b + f_\theta(x)$ to y. The residual and direct formulations therefore share the same action reference; their output parameterizations and initial actions differ.

4.2 Balanced Zero-Residual Supervision

Because the head runs at every query, shifted-view targets alone would leave its clean-input behavior unconstrained. We therefore use four record categories: altered views with the target in Equation 3, their paired clean views, ordinary clean rollouts, and physically shifted rollouts. The last three have zero-residual labels. For record i, the equivalent action reference is $y_i = b_i + \delta_i$; a zero residual asks to retain b_i, not to stop the robot. With inputs x_i and category $\mathcal{D}_c$, we minimize

$$\mathcal{L}(\theta) = \sum_{c=1}^{4} \lambda_c \sum_{i \in \mathcal{D}_c} w_i \frac{\|f_\theta(x_i) - \delta_i\|_F^2}{3H}, \qquad \lambda_c = \tfrac{1}{4}. \tag{4}$$

Weights sum to one within each category. Task–initial-state blocks receive equal mass, divided equally among source policies and then their recorded frames. This prevents long trajectories from dominating simply by supplying more frames. The three zero-target categories together receive 75% of the nominal weight.

Paired clean records constrain the other side of each pair; ordinary rollouts extend preservation beyond intervention states. Physical-shift records supply zero targets for retaining base actions when

the scene changes physically. Category-removal controls keep record order and update budget, zero the removed weights, and rebalance the remaining categories.

4.3 Correction Head and Execution

At each base query, the head predicts

$$r_t = f_\theta(I_t^f, I_t^w, s_t, N(A_t)) \in \mathbb{R}^{H\times 3}. \tag{5}$$

Two 224×224 images are resized to 64×64 and encoded by a shared three-layer CNN. The image features, robot state, and normalized action chunk pass through two width-256 layers to produce $H \times 3$ corrections. The $\pi_{0.5}$ head has 249,150 parameters; its zero-initialized output layer initially preserves the base proposal.

Let $\sigma_k = (q_{0.99,k} - q_{0.01,k} + \epsilon)/2$. We restore action units and apply

$$\widetilde{A}_{t,j,k} = \begin{cases} A_{t,j,k} + g\sigma_k r_{t,j,k}, & k \in \{x, y, z\}, \\ A_{t,j,k}, & \text{otherwise.} \end{cases} \tag{6}$$

For $\pi_{0.5}$, $H = 10$, $h = 5$, and $g = 1$: loss covers ten translations, and five are executed. Separate heads use the native horizons of π_0 ($H = 50$, $h = 5$) and OpenVLA ($H = h = 1$), with the same gain (Appendix P). Correction runs at every query, without clipping, detection, or history. Other coordinates are copied from A_t; each query uses one base call and one head call.

5 Experimental Setup

Tasks and interventions. We evaluate frozen VLA policies on the ten LIBERO-Spatial tasks, using $\pi_{0.5}$ as the primary backbone and π_0 and OpenVLA for replication. Near placement, the target's rendered position is temporarily shifted; non-privileged policies receive no trigger signal. Episodes end at success or after 220 control steps. Failures, untriggered interventions, and interventions producing no pixel change are all retained. Table 1 displays 4/8 cm offsets and 10/20/40 control steps. These conditions were selected for presentation after the original grid was evaluated; intervals retain the original 36-comparison adjustment. The extension increases the duration of 8 cm perturbations to as many as 80 steps (Figure 3). Duration denotes the scheduled persistence of the visual error, not the prediction horizon.

For the primary $\pi_{0.5}$ panels, a benchmark-only trigger checks target geometry, contact, and executed gripper history every five control steps. The initial-state seed fixes the displacement direction. A separate simulator renders the shifted target without moving the target or robot in the active simulator at the moment of intervention. When the scheduled window ends, ordinary rendering resumes from the state reached under the intervening actions. An episode can terminate before the full window is observed.

Training and controls. The principal $\pi_{0.5}$ heads share 2,601 fitting records, inputs, normalization, seed 20260914, and a budget of 50 epochs and 2,050 updates. Direct prediction removes the fixed output skip connection. The mismatched-reference control breaks the correspondence of clean targets within each task, initial state, and source trajectory. Uniform-record weighting removes both category- and block-level balancing. Two additional heads learn a diagonal skip connection initialized to either the identity or zero matrix. Duration-extension tests reuse frozen correction heads, including separately trained π_0 and OpenVLA heads, without further training or configuration selection on the new initial states.

Fitting uses supplied initial-state IDs 0–2 in each task, with IDs 3–4 reserved for development. Final-epoch checkpoints are fixed before native-reset and duration-extension evaluation. The direct head matches PRA's network size and inputs, but its zero-initialized output maps to the action-normalization center instead of initially preserving the base chunk. The residual/direct contrast therefore changes the starting action as well as the output parameterization.

Evaluation and reporting. Original panels A–C each use 300 native-reset initial states; grid and mechanism subsets each use 100. Their 19,400 formal evaluation rollouts reuse 900 generated states.

Table 1: **PRA versus the base policy across offsets and durations.** All methods and conditions use the same 100 initial states from panel A. Success rates are percentages; PRA-minus-base differences and intervals are in percentage points (pp). Intervals retain the original 36-comparison correction, giving approximately 99.861% confidence. Bold denotes PRA. Further displayed 4/8 cm comparisons are in Appendix A; 300-state panels are in Appendix O.

Offset	Steps	Base (%)	**PRA (%)**	Gain (pp)	Adjusted CI (pp)
4 cm	10	98.0	**88.0**	-10.0	$[-20.0, -2.0]$
4 cm	20	86.0	**85.0**	-1.0	$[-15.0, 13.0]$
4 cm	40	78.0	**84.0**	$+6.0$	$[-7.3, 20.0]$
8 cm	10	84.0	**78.0**	-6.0	$[-19.0, 8.0]$
8 cm	20	66.0	**65.0**	-1.0	$[-18.0, 16.0]$
8 cm	40	44.0	**64.0**	$+20.0$	$[0.0, 39.0]$

The duration extension adds separate panels of 300 initial states for $\pi_{0.5}$, 100 for π_0, and 100 for OpenVLA, yielding 6,000, 800, and 800 rollouts, respectively. Methods and conditions share initial states within each panel. Task-stratified paired-bootstrap intervals retain the registered comparison families, and panels are not pooled for inference. A confirmed improvement requires a gain of at least three percentage points and a corrected interval with a strictly positive lower bound. Panels A–C target different questions: A compares residual and direct heads, B examines skip initialization and timing, and C tests pairing and weighting. The grid and timing analyses reuse fixed subsets of A and B, so neither is an independent replication; duration tests use new starts and frozen checkpoints. Full protocols are in Appendices N and P.

The bootstrap resamples paired initial states within each task 50,000 times and gives the ten tasks equal weight. Its intervals describe initial-state variation under fixed tasks and frozen checkpoints, not variation across training seeds. In the duration tests, all assigned episodes remain in the denominator even if the trigger does not fire or the rollout ends before the scheduled perturbation window finishes.

6 Results

6.1 Effect of Perturbation Duration

Table 1 reports the 4/8 cm grid on a 100-state subset of panel A; its 8 cm/20-step cell reuses the main evaluation rollouts. Results for the full 300-state panels are in Appendix O. Figure 3 extends the duration of the 8 cm perturbation on fresh initial states using frozen checkpoints.

At 8 cm, base success falls from 84% to 66% to 44% as the shift lasts 10, 20, and 40 steps; PRA achieves 78%, 65%, and 64%. From 20 to 40 steps, the base loses 22 percentage points and PRA loses one. The $+20$-point estimate at 8 cm/40 steps motivates the independent fresh-state test below; its grid-adjusted interval is $[0, 39]$. At 4 cm/10 steps, PRA instead reduces success from 98/100 to 88/100 (-10 points; original 36-comparison adjusted interval $[-20, -2]$). At 8 cm/20 steps, full-panel gains over the base are 3.7, 6.7, and 3.0 points in A–C; all adjusted intervals include zero. Panel A's 100-state grid subset has one fewer PRA success (65 versus 66), against eleven more on the full 300 states (206 versus 195). The subset is descriptive; the longer-duration test uses new starts. Further interpretation is in Appendices L.1 and O.

Registered duration extension. On 300 new initial states for $\pi_{0.5}$, PRA succeeds in 191/300 episodes at 40 steps, compared with 141/300 for the base: a gain of 16.7 percentage points, with a 97.5% interval of $[8.7, 24.7]$. The gain at 40 steps exceeds that at 20 steps by 13.0 points, with an interval of $[4.3, 21.7]$. Both registered primary comparisons meet the improvement criterion. At 60 and 80 steps, gains are 28.3 and 38.0 points, with 99.375% intervals of $[19.3, 37.3]$ and $[29.3, 46.7]$ in the secondary comparison family. Nine of the ten tasks have positive 40-step point-estimate gains (Appendix P).

Replication with π_0. On a separate 100-state panel, PRA raises 40-step success from 49% to 62%, a gain of 13 points with a 97.5% interval of $[1, 25]$. The registered secondary comparison at 80 steps

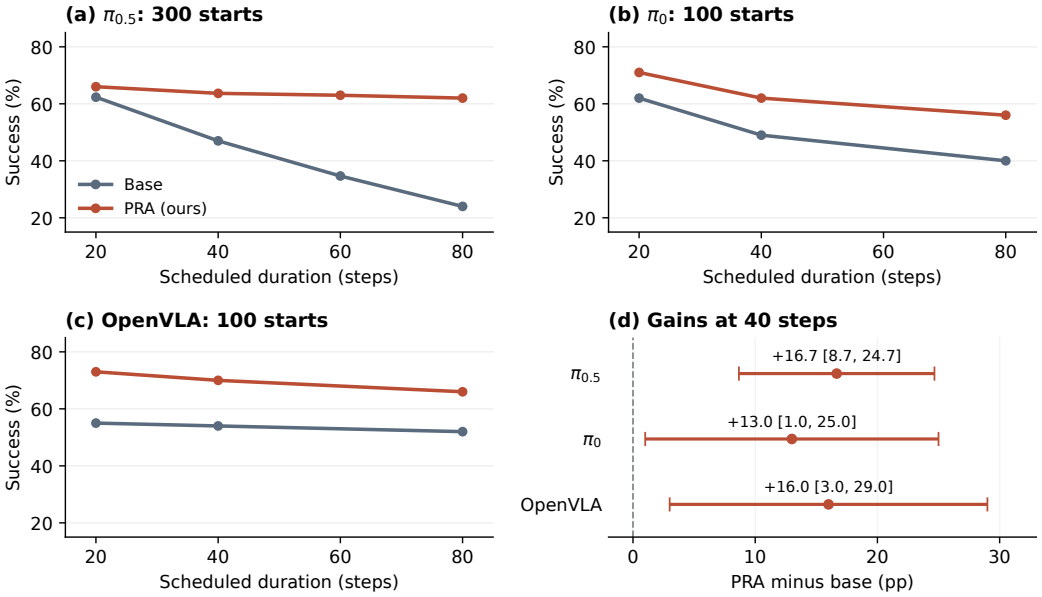


Figure 3: **Longer visual perturbations across three backbones.** (a–c) Success under 8 cm offsets, evaluated on separate fresh-state panels of 300 starts for $\pi_{0.5}$, 100 for π_0, and 100 for OpenVLA. (d) PRA-minus-base gains at 40 steps, with 97.5% paired intervals for each experiment's two-primary-comparison family. The π backbones execute five actions per query, while OpenVLA executes one. Durations are scheduled values; realized exposure varies across rollouts. Clean results, interaction tests, and complete comparisons are in Appendix P.

yields a 16-point gain, with a 95% interval of $[5, 27]$. Seven of ten tasks have positive 40-step point-estimate gains and the other three tie. Complete conditions, interactions, and exposure statistics are in Appendix P.

Replication with OpenVLA. On another 100 initial states, PRA raises 40-step success from 54% to 70%, a gain of 16 points with a 97.5% interval of $[3, 29]$. The registered secondary comparison at 80 steps gives a 14-point gain, with a 95% interval of $[3, 24]$. Seven tasks have positive 40-step point-estimate gains and three tie. This extends the 40-step benefit to a single-action backbone, using the full PRA extension with wrist-image and state inputs. A widening 20-to-40-step advantage is confirmed only for $\pi_{0.5}$; all interaction tests are in Appendix P.

At 40 steps, PRA recovers 90 base failures and loses 40 base successes with $\pi_{0.5}$; the corresponding counts are 22 and nine with π_0, and 27 and 11 with OpenVLA. These are paired outcome changes within separate panels, not pooled observations. The 20-to-40-step interaction is uncertain for π_0 ($+4$ points, 97.5% interval $[-9, 17]$) and OpenVLA (-2 points, $[-15, 11]$), so the widening advantage observed for $\pi_{0.5}$ does not extend to either replication.

Scheduled duration differs from realized exposure. In the $\pi_{0.5}$ 40-step test, the base and PRA experience a mean of 38.92 and 27.65 shifted steps, respectively; the complete window occurs in 288/300 and 148/300 episodes. Because termination depends on the policy's trajectory, restricting analysis to complete windows would change the tested population. The registered comparisons retain all assigned starts (Appendix P).

6.2 Design Comparisons under Matched Budgets

Pairing and weighting. In panel C's 8 cm/20-step condition, PRA succeeds in 217/300 episodes, compared with 201/300 for mismatched references and 218/300 for uniform-record weighting (Table 2). These matched-budget controls delimit what can be attributed to the supervision design; the complete native-reset results are in Appendix O.

Table 2: **Pairing and weighting controls in panel C.** Success rates are percentages over 300 paired initial states per condition. The last column reports PRA minus the policy in that row under 8 cm/20-step perturbations, in percentage points, with 98.333% corrected intervals. Uniform-record weighting retains paired targets but removes category- and block-level balancing. Transfer results are in Appendix K.

Policy	Clean	Visual	PRA − row [98.333% CI]
Base policy	97.7	69.3	$+3.0$ $[-5.0, 11.0]$
PRA (ours)	**91.3**	**72.3**	—
Mismatched references	93.0	67.0	$+5.3$ $[-2.0, 12.7]$
Record-uniform training	91.3	72.7	-0.3 $[-6.3, 5.7]$

Table 3: **Parameterization and initialization on panel B (300 starts).** Visual: 8 cm/20 steps. The learned-skip pair shares an architecture. Bottom: five primary differences with 99% adjusted intervals; † marks a confirmed gain.

Policy	Initial skip D	Clean	Visual
Base policy	—	292/300	195/300
PRA (fixed residual)	$\boldsymbol{I}$ **(fixed)**	**275/300**	**215/300**
Direct prediction	0 (fixed)	266/300	176/300
Learnable skip, identity init.	I (learned)	272/300	211/300
Learnable skip, zero init.	0 (learned)	256/300	181/300

Primary comparison	Difference [99% CI] (pp)
Learnable identity − learnable zero	$+10.0$ $[1.3, 18.7]^{\dagger}$
PRA − direct prediction	$+13.0$ $[5.3, 20.7]^{\dagger}$
Learnable identity − **PRA**	-1.3 $[-8.0, 5.3]$
Learnable identity − base	$+5.3$ $[-3.3, 14.0]$
PRA − base	$+6.7$ $[-2.3, 15.7]$

These comparisons do not isolate a closed-loop gain from pairing or balancing alone. Longer-duration contrasts and training-target diagnostics are reported in Appendices P and Q.

Residual versus direct prediction. PRA exceeds direct prediction by 9.3 percentage points in panel A, with a corrected interval of $[2.0, 16.7]$, and by 13.0 points in panel B, with an interval of $[5.3, 20.7]$. Both heads receive the base action chunk as input. The residual formulation also retains it at the output, so this comparison changes both the fixed skip connection and the initial action.

Initialization within one architecture. For the model $Db + f_\theta(x)$, initializing the learnable D to the identity yields 211/300 visual successes, compared with 181/300 for zero initialization: a 10-point difference with a 99% interval of $[1.3, 18.7]$ (Table 3). Architecture and parameter count are identical, isolating the initialization choice within this model family. In panel B, these design contrasts favor residual prediction and identity initialization, while PRA's adjusted interval against the base still includes zero; neither contrast alone establishes a 20-step gain. Offline training diagnostics are in Appendix I.

6.3 Execution Timing and Reference Diagnostics

Table 4 compares execution rules on the same 100 mechanism states from panel B under the 8 cm condition. Always-on PRA succeeds in 80 episodes using no event timing; window-only correction succeeds in 81 with privileged start and end times. Their one-point difference has a corrected interval of $[-10, 11]$. Full timing comparisons are in Appendix K.

Using clean-reference XYZ actions during the visual window succeeds in 99/100 episodes, exceeding learned window-only correction by 18 percentage points, with a corrected interval of $[7, 29]$. This privileged-reference diagnostic indicates room for stronger action correction; it requires clean images and extra base-model queries (Appendix L).

Table 4: **Timing and reference diagnostics (panel B, 8 cm, 100 starts).** Entries are success counts. Timing and clean images are privileged; "None" means neither privileged input nor extra base query. Intervals retain the original ten-comparison correction (99.5%; Appendix K).

Controller	Privileged inputs / extra queries	Successes
Base policy	None	70/100
PRA (all-time)	**None**	**80/100**
Window-only residual	Start and end times	81/100
Trigger-onward residual	Trigger time	72/100
Clean-reference window	Timing, clean images, extra queries	99/100

Comparison	Difference [99.5% CI] (pp)
Window-only − base	$+11.0$ $[-3.0, 25.0]$
Trigger-onward − window-only	-9.0 $[-19.0, 0.0]$
Clean-reference − window-only	$+18.0$ $[7.0, 29.0]$
Window-only − **PRA**	$+1.0$ $[-10.0, 11.0]$
Clean-reference − base	$+29.0$ $[16.0, 42.0]$

PRA also changes behavior outside the intended visual-shift condition. In panel C, clean success is 274/300 with PRA and 293/300 with the base policy (-6.3 points; descriptive 95% interval $[-9.3, -3.7]$). The fresh $\pi_{0.5}$ duration panel shows a similar clean loss, 275/300 versus 293/300; π_0 and OpenVLA have equal aggregate clean counts, which does not establish equivalence. When the target physically moves in panel B, the counts are 63/300 and 254/300 for PRA and the base, respectively. These checks use separate state panels and are not pooled with the visual-shift comparisons (Appendix L).

7 Discussion and Conclusion

PRA is most effective in the tested regime where the target appears displaced near placement but remains physically fixed. Fixed residual heads improve 40-step success on new initial states with $\pi_{0.5}$, π_0, and OpenVLA. For $\pi_{0.5}$, the improvement over the base policy also grows from 20 to 40 steps. Under matched training budgets, correcting the base proposal works better than predicting actions directly, and identity initialization helps when the skip connection is learned. These comparisons locate the benefit in a small action correction applied at the policy's normal query cadence.

The head runs at every query, whether or not the base action needs correction. It receives neither event timing nor observation history that could help separate an apparent shift from a target that really moved. The clean-view action used for training is another prediction from the frozen policy, not an expert recovery action. The experiments do not isolate the cause of each failure. They show a narrower pattern: PRA helps with the tested longer rendering-only shifts, but can hurt on short shifts, clean episodes, or physical target movement.

Paired predictions from a frozen VLA provide the training signal without additional expert action labels. At deployment, the residual head needs neither a clean reference image nor extra base-model queries. In the tested LIBERO-Spatial setting, this offers action adaptation for transient visual mislocalization, with the detailed boundaries and clean-condition results in Appendices L–M.

Ethics Statement

All experiments are conducted in simulation and do not establish safe deployment on physical robots. Visual errors and physical changes may coexist, so deployment requires environment-specific validation, monitoring, and failure handling. We retain adverse conditions and the complete prespecified sample denominators, and do not treat success on visual-error cases as evidence of general safety.

Statement on the Use of Generative AI

OpenAI Codex was used for language editing and polishing; refinement of mathematical notation and descriptions; visual and formatting improvements to tables and figures; LaTeX formatting and typesetting; and assistance with generating perturbations. Perturbed images were rendered in the LIBERO/MuJoCo simulation environment according to explicitly specified displacement, duration, and random-seed rules. No generative image or video models were used to synthesize experimental imagery. Codex was not used to design the methodology or experiments, generate or analyze experimental data, or determine experimental outcomes or scientific conclusions. Numerical results, tables, and figures are based on archived experimental reports. The authors reviewed the AI-assisted edits and take full responsibility for the manuscript, its conclusions, and the supporting evidence.

Reproducibility Statement

Sections 3–5 define the inputs, targets, weighting, execution, and evaluation. Appendices C, N, and P provide implementation and protocol details; Appendices A and K provide the displayed grid comparisons and further panel controls. Development results, proxy diagnostics, physical-target-displacement results, and provenance records are presented separately. Appendix J documents recovery of the diagnostic experiments. Reconstructing archived reports is distinct from rerunning policy inference or independently relabeling success from videos.

REFERENCES

Christopher Agia, Rohan Sinha, Jingyun Yang, Ziang Cao, Rika Antonova, Marco Pavone, and Jeannette Bohg. Unpacking failure modes of generative policies: Runtime monitoring of consistency and progress. In *Proceedings of The 8th Conference on Robot Learning*, volume 270, pp. 689–723. PMLR, 2025. URL https://proceedings.mlr.press/v270/agia25a.html.

Lars Ankile, Anthony Simeonov, Idan Shenfeld, Marcel Torne, and Pulkit Agrawal. From imitation to refinement – residual RL for precise assembly. *arXiv preprint arXiv:2407.16677*, 2024. URL https://arxiv.org/abs/2407.16677.

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 *Proceedings of Robotics: Science and Systems*, 2025a. doi: 10.15607/RSS.2025.XXI.010. URL https://www.roboticsproceedings.org/rss21/p010.html.

Kevin Black, Manuel Galliker, and Sergey Levine. Real-time execution of action chunking flow policies. In *Advances in Neural Information Processing Systems*, volume 38, 2025b. doi: 10.52202/085713-1122. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/300ccb2187dedd4edcc07f7e76d8e553-Abstract-Conference.html.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin CM Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.026. URL https://www.roboticsproceedings.org/rss19/p026.html.

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.

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. On robustness of vision-language-action model against multi-modal perturbations. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/72802bef5cf1a3449e909b20c2ae18d5-Abstract-Conference.html.

Bear Häon, Kaylene Caswell Stocking, Ian Chuang, and Claire Tomlin. Mechanistic interpretability for steering vision-language-action models. In *Proceedings of The 9th Conference on Robot Learning*, volume 305, pp. 2743–2762. PMLR, 2025. URL https://proceedings.mlr.press/v305/haon25a.html.

Bingqi Huang, Bingchuan Wei, Xuan Wang, Yingkai Cai, and Zhaokui Wang. Cross-view action consistency for camera-robust vision-language-action policies. *arXiv preprint arXiv:2608.06965*, 2026. URL https://arxiv.org/abs/2608.06965.

Tobias Johannink, Shikhar Bahl, Ashvin Nair, Jianlan Luo, Avinash Kumar, Matthias Loskyll, Juan Aparicio Ojea, Eugen Solowjow, and Sergey Levine. Residual reinforcement learning for robot control. *arXiv preprint arXiv:1812.03201*, 2018. doi: 10.48550/arXiv.1812.03201. URL https://arxiv.org/abs/1812.03201.

Moo Jin Kim, Chelsea Finn, and Percy Liang. Fine-tuning vision-language-action models: Optimizing speed and success. In *Proceedings of Robotics: Science and Systems*, 2025a. doi: 10.15607/RSS.2025.XXI.017. URL https://www.roboticsproceedings.org/rss21/p017.html.

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan P Foster, Pannag R Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. OpenVLA: An open-source vision-language-action model. In *Proceedings of The 8th Conference on Robot Learning*, volume 270, pp. 2679–2713. PMLR, 2025b. URL https://proceedings.mlr.press/v270/kim25c.html.

Jacky Kwok, Christopher Agia, Rohan Sinha, Matt Foutter, Shulu Li, Ion Stoica, Azalia Mirhoseini, and Marco Pavone. RoboMonkey: Scaling test-time sampling and verification for vision-language-action models. In *Proceedings of The 9th Conference on Robot Learning*, volume 305, pp. 3200–3217. PMLR, 2025. URL https://proceedings.mlr.press/v305/kwok25a.html.

Michael Laskey, Jonathan Lee, Roy Fox, Anca Dragan, and Ken Goldberg. DART: Noise injection for robust imitation learning. In *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78, pp. 143–156. PMLR, 2017. URL https://proceedings.mlr.press/v78/laskey17a.html.

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*, volume 36, 2023. doi: 10.52202/075280-1939. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/8c3c666820ea055a77726d66fc7d447f-Abstract-Datasets_and_Benchmarks.html.

Yuejiang Liu, Jubayer Hamid, Annie Xie, Yoonho Lee, Max Du, and Chelsea Finn. Bidirectional decoding: Improving action chunking via guided test-time sampling. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/0d78dd998f7b9ac79604d47a2d79bb0d-Abstract-Conference.html.

Jingzhou Luo, Yifan Wen, Yongjie Bai, Xinshuai Song, Yang Liu, and Liang Lin. RoVLA: Multi-consistency constraints for robust vision-language-action models. *arXiv preprint arXiv:2605.19678*, 2026. URL https://arxiv.org/abs/2605.19678.

Karl Pertsch, Kyle Stachowicz, Brian Ichter, Danny Driess, Suraj Nair, Quan Vuong, Oier Mees, Chelsea Finn, and Sergey Levine. FAST: Efficient action tokenization for vision-language-action models. In *Proceedings of Robotics: Science and Systems*, 2025. doi: 10.15607/RSS.2025.XXI.012. URL https://www.roboticsproceedings.org/rss21/p012.html.

Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Manuel Y. Galliker, Dibya Ghosh, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Devin LeBlanc, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Allen Z. Ren, Lucy Xiaoyang Shi, Laura Smith, Jost Tobias Springenberg, Kyle Stachowicz, James Tanner, Quan Vuong, Homer Walke, Anna Walling, Haohuan Wang, Lili Yu, and Ury Zhilinsky. $\pi_{0.5}$: A vision-language-action model with open-world generalization. *arXiv preprint arXiv:2504.16054*, 2025. URL https://arxiv.org/abs/2504.16054.

Stephane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, pp. 627–635. PMLR, 2011. URL https://proceedings.mlr.press/v15/ross11a.html.

Kohei Sendai, Maxime Alvarez, Tatsuya Matsushima, Yutaka Matsuo, and Yusuke Iwasawa. Leave no observation behind: Real-time correction for VLA action chunks. *arXiv preprint arXiv:2509.23224*, 2025. URL https://arxiv.org/abs/2509.23224.

Tom Silver, Kelsey Allen, Josh Tenenbaum, and Leslie Kaelbling. Residual policy learning. *arXiv preprint arXiv:1812.06298*, 2018. doi: 10.48550/arXiv.1812.06298. URL https://arxiv.org/abs/1812.06298.

Junhyuk So, Chiwoong Lee, Shinyoung Lee, Jungseul Ok, and Eunhyeok Park. Improving generative behavior cloning via self-guidance and adaptive chunking. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-2830. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/79ce24f9e8d3c4ff5919240eac78a782-Abstract-Conference.html.

Yihao Wang, Pengxiang Ding, Lingxiao Li, Can Cui, Zirui Ge, Xinyang Tong, Wenxuan Song, Han Zhao, Wei Zhao, Pengxu Hou, Siteng Huang, Yifan Tang, Wenhui Wang, Ru Zhang, Jianyi Liu, and Donglin Wang. VLA-Adapter: An effective paradigm for tiny-scale vision-language-action model. *arXiv preprint arXiv:2509.09372*, 2025. URL https://arxiv.org/abs/2509.09372.

Andrew Warrington, Jonathan W Lavington, Adam Scibior, Mark Schmidt, and Frank Wood. Robust asymmetric learning in POMDPs. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pp. 11013–11023. PMLR, 2021. URL https://proceedings.mlr.press/v139/warrington21a.html.

Chuyan Xiong, Chengyu Shen, Xiaoqi Li, Kaichen Zhou, Jiaming Liu, Ruiping Wang, and Hao Dong. Autonomous interactive correction MLLM for robust robotic manipulation. In *Proceedings of The 8th Conference on Robot Learning*, volume 270, pp. 3139–3156. PMLR, 2025. URL https://proceedings.mlr.press/v270/xiong25a.html.

Xiu Yuan, Tongzhou Mu, Stone Tao, Yunhao Fang, Mengke Zhang, and Hao Su. Policy decorator: Model-agnostic online refinement for large policy model. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/45c361d4117d598d4bb6568b407e9ac9-Abstract-Conference.html.

Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. In *Proceedings of Robotics: Science and Systems*, 2023. doi: 10.15607/RSS.2023.XIX.016. URL https://www.roboticsproceedings.org/rss19/p016.html.