CAUSAL NUISANCE INTERVENTIONS FOR DISTRACTOR-ROBUST WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

World models trained from pixels can remain sensitive to camera motion and changing backgrounds, even when trained without pixel reconstruction. We introduce *Causal Nuisance Interventions* (CaNI), a causal formulation and replay-training method using a specified nuisance structural causal model (SCM) to guide observation-proxy design before training. We instantiate CaNI on a decoder-free Dreamer backbone by applying one randomly sampled padded translation consistently across each replay sequence, while retaining the architecture and learning objectives. We prove that uniform posterior stability is sufficient to bound changes in a fixed recurrent policy's return under observation interventions that preserve the controlled process. Across five DeepMind Control tasks with driving-video backgrounds and camera jitter, CaNI achieves higher mean return than each of five evaluated baselines on every task, increasing task-averaged return from 310.5 for the matched backbone to 715.5 for CaNI (approximately 130%). On the driving-video benchmark without jitter, time-averaged return improves by 15% with comparable final returns. Frozen-policy evaluations show improved robustness to evaluation-time jitter, and disjoint training and evaluation splits of the DAVIS video dataset show better performance on unseen backgrounds on two tasks. Posterior diagnostics and controlled ablations support reduced shift sensitivity and the intervention design.

1 INTRODUCTION

World models trained from pixels use inferred latent states to learn dynamics and initialize imagined control (Hafner et al., 2025; Morihira et al., 2026). Yet pixels can change even when the controlled system does not: backgrounds vary across settings, and camera displacement moves task-relevant content within the image (Tobin et al., 2017; Stone et al., 2021). If the latent-state posterior responds to this variation, nuisance information can propagate through recurrent memory, imagined dynamics, and action selection (Wang et al., 2022; Fu et al., 2021). Decoder-free objectives remove the pressure to reconstruct irrelevant pixels. Figure 1 compares learning under video backgrounds and camera displacement, alongside an empirical extension to corrupted reward measurements: decoder-free R2-Dreamer outperforms reconstruction-based DreamerV3 in the two video settings, but its return drops when camera jitter is added. This sensitivity motivates examining how nuisance mechanisms enter representation learning and control.

Existing methods improve visual robustness through image augmentation (Laskin et al., 2020b; Yarats et al., 2021; Hansen & Wang, 2021), representation objectives, or structured latent models (Deng et al., 2022; Liu et al., 2023). DreamerPro already applies temporally consistent random shifts before recurrent world-model inference (Deng et al., 2022). Linking this augmentation to a nuisance mechanism requires a proxy that approximates its effect while preserving the physical process, control targets, and information needed for control. Image-plane displacement motivates translation; changing an occluded background generally requires segmentation or renderer access.

We introduce *Causal Nuisance Interventions* (CaNI), which uses a manually specified nuisance structural causal model (SCM) (Pearl, 2009) to guide observation-proxy design before training. Our implementation applies one randomly sampled padded translation to every frame of a replay sequence before world-model inference on a decoder-free Dreamer backbone (Yarats et al., 2021; Deng et al., 2022).

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.

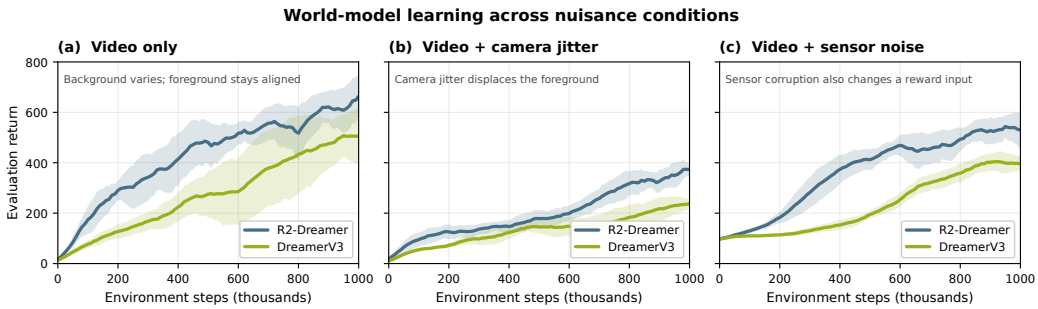

Figure 1: Cheetah Run with driving-video backgrounds: (a,b) observation nuisances; (c) reward-sensor noise with perturbed training/evaluation rewards. Curves show smoothed three-seed means; bands show sample standard deviations.

The SCM identifies which mechanism to vary and which control quantities to preserve. Training samples from the fixed proxy family, retaining the architecture, objectives, and recorded control targets. We use the same translation rule across all evaluated nuisance settings. Acting and evaluation use environment observations without additional CaNI augmentation. Figure 2 summarizes the path from nuisance structure to training.

We ask what stability property of a recurrent world model is sufficient for policy robustness when a visual transformation preserves the underlying control process. We formalize visual nuisances as changes to the observation mechanism that preserve the joint controlled transition, reward, and continuation kernel. Under this formulation, a uniform bound on posterior sensitivity quantitatively bounds the change in a recurrent policy's return across supported nuisance-intervention schedules. At exact posterior stability, an additional control-sufficiency condition yields an invariant Markov decision process (MDP) abstraction that preserves optimal policies (Gelada et al., 2019; Zhang et al., 2021).

Across five visual-control tasks, CaNI improves sample efficiency with video backgrounds and achieves higher mean returns than all five evaluated baselines under camera jitter. Frozen-policy evaluations show greater robustness to evaluation-time jitter, and held-out DAVIS backgrounds demonstrate generalization to unseen videos. In the reward-sensor extension, CaNI improves mean measured return over R2-Dreamer on Cheetah and Hopper, with near ties on Cartpole and Walker. Posterior analysis and controlled ablations connect the visual-robustness findings to reduced sensitivity and the intervention design. Our contributions are:

- a causal observation-channel formulation with a recurrent-control return bound under uniform posterior stability and an exact-abstraction corollary under control sufficiency;
- CaNI, a method for training over SCM-motivated observation proxies, instantiated by sequence-consistent translation on an unchanged decoder-free backbone; and
- evidence of higher control returns under camera jitter, faster learning without jitter, and generalization to unseen backgrounds, supported by frozen-policy tests, posterior diagnostics, and ablations.

2 Causal Formulation of Nuisance-Robust Control

2.1 Recurrent visual control

At decision t, let H_t be the agent's recurrent memory before assimilating observation O_t; lowercase letters denote realizations of random variables. The posterior kernel $\Phi(dz \mid h, o)$ samples a latent representation Z_t in latent space $\mathcal{Z}$. A policy kernel π selects action A_t from the learned state S_t, and a deterministic map F updates memory:

$$Z_t \sim \Phi(\cdot \mid H_t, O_t), \qquad S_t = (H_t, Z_t), \qquad A_t \sim \pi(\cdot \mid S_t), \qquad H_{t+1} = F(H_t, Z_t, A_t). \quad (1)$$

The learned state S_t supports prediction and control. We use a decoder-free categorical recurrent state-space model (RSSM); Appendix B specifies its architecture and objective.

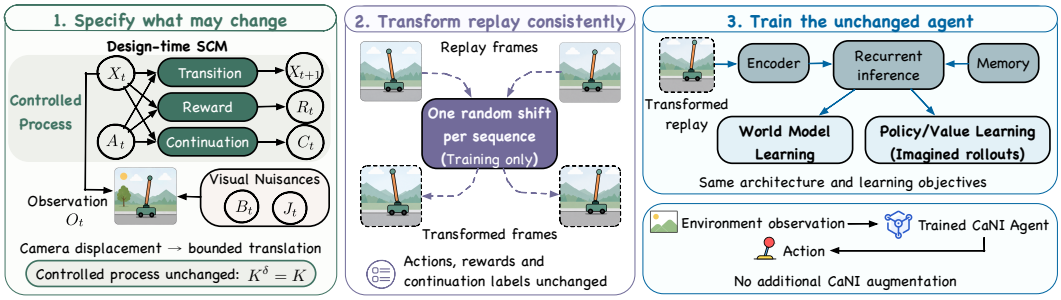

Figure 2: CaNI: design-time SCM guidance, sequence-shared replay translation, and training with unchanged architecture and objectives. X_t, R_t, C_t denote physical state, reward and continuation; B_t, J_t denote background and camera nuisances. The joint transition, reward and continuation kernel K is preserved under translation with offset δ: $K^\delta = K$.

Assumption 1 (Nuisance-stable recurrent update)**.** *The initial memory H_0 is generated by a nuisance-independent rule, and the recurrent update has the form in Equation 1. Thus the current observation can affect future memory only through the realized posterior draw Z_t.*

2.2 Causal Nuisance Interventions

We express visual nuisances through background and camera variables, then analyze supported observation channels that preserve the controlled process. Let X_t be the task-relevant physical state, R_t the reward, C_t the continuation (non-termination) signal, and B_t and J_t the background and camera variables. A base renderer $\mathcal{R}_0$ forms a scene image from the physical state and background; a camera map G_{J_t} then transforms it. Let K denote the joint controlled kernel for next physical state, reward, and continuation.

Assumption 2 (Visual nuisances are not task causes)**.**

$$(X_{t+1}, R_t, C_t) \sim K(\cdot \mid X_t, A_t), \tag{2}$$

$$O_t = G_{J_t}\big(\mathcal{R}_0(X_t, B_t)\big). \tag{3}$$

The nuisance variables enter only the observation stages and leave the joint controlled kernel K unchanged given (X_t, A_t). A time-limit clock is included in X_t.

Interventions on B_t or J_t therefore change only the observation. We formalize the target as a family $\mathcal{I}$ of *controlled-process-preserving observation channels*: alternative observation rules that leave K unchanged. The family may contain renderer-level background or camera changes, or image-space channels such as CaNI's translation (Section 3). A *causal intervention schedule* selects at each time a channel in $\mathcal{I}$ using only variables determined before the current observation is delivered and changes only the observation mechanism. For channels $\iota, \iota' \in \mathcal{I}$, paired observations $o^\iota, o^{\iota'}$ fix the physical state and all nonintervened variables and exogenous draws.

2.3 Posterior Stability and Control Robustness

We fix the trained posterior Φ, recurrent update F, and policy π while varying the observation-intervention schedule. At a shared physical state and recurrent memory, two runs receive different nuisance-transformed observations: close posteriors allow their latent draws to be coupled to agree with high probability. Shared policy and controlled-process randomness propagates agreement through actions and memory, limiting return differences.

For probability measures μ, ν on $\mathcal{Z}$, total variation (TV) is $\mathrm{TV}(\mu, \nu) := \sup_E |\mu(E) - \nu(E)|$, where the supremum is over measurable sets $E \subseteq \mathcal{Z}$. For a causal intervention schedule Q supported on $\mathcal{I}$ and recurrent policy π, write $\mathbb{E}^\pi_Q$ for expectation under Q and π. With discount factor $\gamma \in (0, 1)$, the expected discounted return is $J_Q(\pi) = \mathbb{E}^\pi_Q[\sum_{t=0}^\infty \gamma^t R_t]$.

Theorem 1 (Causal nuisance robustness of recurrent control)**.** *Suppose Assumptions 1 and 2 hold and the latent space $\mathcal{Z}$ is standard Borel. Let $R_{\max}$ bound rewards as $0 \le R_t \le R_{\max}$, with zero reward after termination. Suppose a uniform tolerance $\varepsilon \in [0,1]$ satisfies, for every pair (h,x) jointly reachable under any supported intervention schedule and any recurrent policy, every pair of channels $\iota, \iota' \in \mathcal{I}$, and every corresponding paired observation $o^{\iota}, o^{\iota'}$,*

$$\mathrm{TV}\Big(\Phi(\cdot \mid h, o^{\iota}), \Phi(\cdot \mid h, o^{\iota'})\Big) \le \varepsilon. \tag{4}$$

Then, for any causal intervention schedules Q, Q' supported on $\mathcal{I}$ with a common initial physical-state law, every fixed recurrent policy π satisfies

$$|J_Q(\pi) - J_{Q'}(\pi)| \le \frac{R_{\max}\varepsilon}{(1-\gamma)^2}. \tag{5}$$

When $\varepsilon = 0$, the joint physical-state, learned-state, action, reward, and continuation trajectory laws are identical under Q and Q'.

Control sufficiency strengthens exact stability to preservation of optimal control.

Corollary 1 (Exact invariant control abstraction)**.** *Under Theorem 1 with $\varepsilon = 0$, let $\mathcal{H}_t$ denote the agent's complete pre-action history, including the current observation and internal draws determining S_t. Assume control sufficiency: a common kernel $\overline{P}$ satisfies, for every history-based policy μ, supported schedule Q, and time t,*

$$\mathrm{Law}^{\mu}_{Q}(R_t, C_t, S_{t+1} \mid \mathcal{H}_t, A_t) = \overline{P}(\cdot \mid S_t, A_t)$$

almost surely on reachable history–action pairs. Here Law^{μ}_{Q} denotes the conditional distribution under μ and Q. Then S_t defines an exact controlled Markov abstraction invariant across intervention schedules. Any optimal policy of this abstract MDP lifts to an optimal history-based policy for the original control process, with the same value under every schedule supported on $\mathcal{I}$.

Proof sketch. Under the coupling above, the probability of a first disagreement at each step, conditional on agreement so far, is at most ε. The disagreement probability by time t is therefore at most $(t+1)\varepsilon$, and summing the discounted reward difference gives Equation 5. At $\varepsilon = 0$, equality persists for the complete trajectory. The additional control-sufficiency condition in Corollary 1 makes the Bellman operator close on S_t, yielding the exact abstraction and optimal-policy lift. Appendix A gives full proofs and a finite-horizon bound for undiscounted returns.

The analysis connects two requirements: the SCM identifies interventions that preserve the controlled process, and posterior stability bounds their effect on recurrent-policy return. Sections 3 and 4.4 develop proxy training and assess posterior sensitivity.

3 CaNI: SCM-Guided Proxy-Intervention Training

For image-plane camera displacement, CaNI applies one padded spatial translation to every frame of a sampled replay sequence, retaining its recorded actions, rewards, and continuation labels. The backbone architecture and learning objectives remain unchanged. The analysis in Section 2 characterizes when posterior stability to such observation changes yields recurrent-policy robustness.

3.1 SCM-guided proxy selection

We specify the nuisance SCM at design time to select proxies using the nuisance's causal location and its action on observations. The location determines which dynamics and labels to preserve; the observation mechanism determines the changes to approximate. In Assumption 2, camera displacement acts through G_{J_t} while leaving K fixed, motivating bounded image translation $O_t^{\delta} = T_{\delta}(O_t)$ (Yarats et al., 2021), with $\delta \sim Q_{\delta}$. Here $\delta = (\delta_x, \delta_y)$ is an integer horizontal/vertical sampling offset in pixels, drawn from the fixed training distribution Q_{δ}, uniform on $\{-4, \dots, 4\}^2$ by default (pad-4). This *camera-aligned translation proxy* defines $\mathcal{I}_{\mathrm{CaNI}} = \{T_{\delta} : \delta \in \mathrm{supp}(Q_{\delta})\}$, where supp denotes distributional support. Each T_{δ} changes only the observation mechanism, preserving the controlled kernel as $K^{\delta} = K$.

Proxy selection combines mechanism alignment, preservation of K, and retention of control information. Bounded translation is appropriate when task-relevant relationships remain observable or recoverable from transformed histories. Exact abstraction in Corollary 1 additionally requires control sufficiency of the learned state. Appendix B.3 specifies the geometry.

3.2 Sequence-level interventional replay

CaNI adds an integer translation to each recorded sequence, using replicate padding to extend boundary pixels beyond the image edges. The sequence may already contain time-varying benchmark jitter. A single displacement is shared across its T frames, following temporally consistent augmentation (Deng et al., 2022). For observations $o_{1:T} = (o_1, \dots, o_T)$, define

$$\mathsf{T}_\delta(o_{1:T}) = \big(T_\delta(o_1), \dots, T_\delta(o_T)\big), \qquad \delta \sim Q_\delta \text{ once per sequence.} \tag{6}$$

The added offset stays constant across the sequence; resampling it at each replay draw exposes the agent to different image locations. Independent per-frame shifts add variation that the model could interpret as physical motion. Let $\mathcal{D}$ denote the replay distribution. For a trajectory $\tau = (o_{1:T}, a_{1:T}, r_{1:T}, c_{1:T}) \sim \mathcal{D}$, the intervened replay is

$$\tau^\delta = \mathsf{T}_\delta(\tau) = \big(\mathsf{T}_\delta(o_{1:T}),\, a_{1:T},\, r_{1:T},\, c_{1:T}\big). \tag{7}$$

3.3 Complete training objective

Let Θ collect all trainable agent parameters and $\mathcal{L}_{\text{agent}}(\tau; \Theta)$ denote its complete world-model and behavior-learning objective on replay sequence τ. CaNI optimizes

$$\min_{\Theta}\ \mathbb{E}_{\tau\sim\mathcal{D},\,\delta\sim Q_\delta}\big[\mathcal{L}_{\text{agent}}\big(\mathsf{T}_\delta(\tau); \Theta\big)\big]. \tag{8}$$

Equation 8 averages the agent objective over intervened replay views. Applying T_δ before encoding makes posterior inference, world-model learning, and imagination starts derive from each transformed sequence, as in augmentation-based world-model training (Deng et al., 2022). Theorem 1 characterizes recurrent-policy robustness under uniform posterior-TV stability. We assess the agent trained by Equation 8 through complementary posterior-sensitivity measurements on held-out histories and complete-rollout robustness tests. CaNI adds no parameters. Acting and evaluation use environment observations without additional CaNI augmentation; Appendix B gives details.

4 Experiments

We organize the experiments around three research questions (RQs). **RQ1** examines control performance, sample efficiency, and robustness to evaluation-time camera changes and unseen backgrounds. **RQ2** asks whether CaNI reduces posterior sensitivity to camera-aligned interventions. **RQ3** studies temporal coupling, intervention family, and translation radius.

4.1 Experimental setup

We evaluate Cartpole Swingup, Hopper Stand, Cheetah Run, Walker Walk, and Cup Catch from the DeepMind Control (DMC) Suite (Tassa et al., 2018). The primary benchmark adds camera jitter to Denoised MDP's driving-video backgrounds (Wang et al., 2022), with a matched no-jitter control. This jitter is applied after rendering as temporally correlated subpixel image translation with bilinear interpolation and black borders. Held-out DAVIS videos (Pont-Tuset et al., 2017) test generalization to unseen backgrounds. Reward-sensor noise on Cartpole, Cheetah, Walker, and Hopper evaluates robustness to corrupted reward measurements.

R2-Dreamer (Morihira et al., 2026) learns decoder-free representations through redundancy reduction and serves as CaNI's matched backbone control. DreamerV3 (Hafner et al., 2025) learns a reconstruction-based world model for imagined policy training. DreamerPro (Deng et al., 2022) replaces pixel reconstruction with prototype prediction across temporally consistent augmented views. iFactor (Liu et al., 2023) factors latent states by action influence and reward relevance. CsDreamer (Zhang et al., 2025) uses a collider-structured RSSM and decoupling regularization to model task-relevant and nuisance factors. Appendix B.1 provides further descriptions.

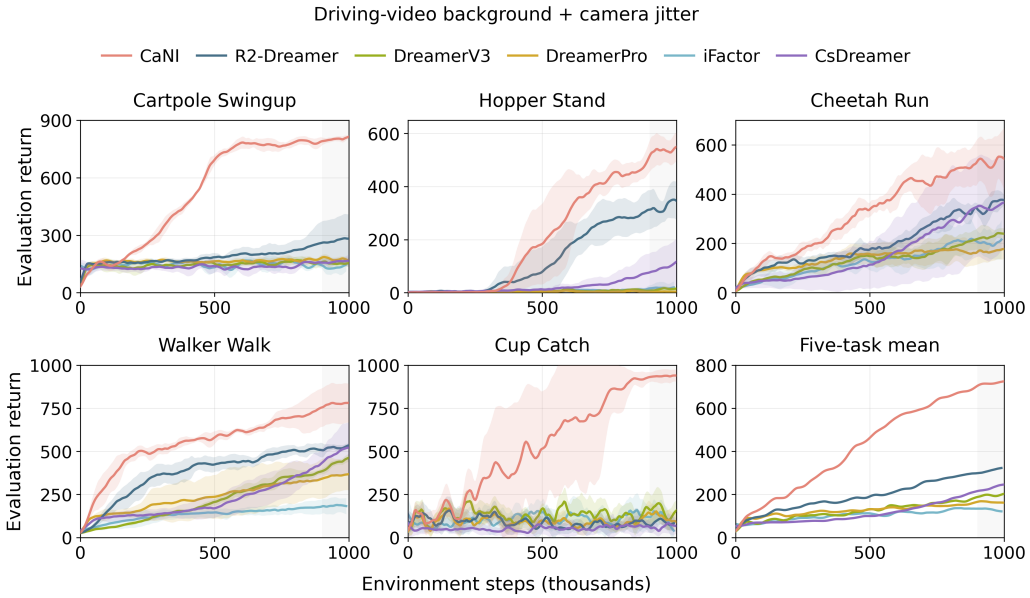

Figure 3: Driving-video backgrounds with camera jitter: three-seed means and sample standard deviations. The last panel averages five tasks equally for all six methods.

Driving-video training and evaluation share a 12-video pool; DAVIS uses 60 training and 30 disjoint evaluation videos. Learning curves use each method's reported step counter through 1M steps (Appendix B.4). Final-window scores average ten raw, ten-episode evaluations at 900k–990k steps, then across seeds; frozen-policy tests use exact-1M checkpoints. Task means receive equal weight. Each condition uses three training seeds; bands and error bars show sample standard deviations (SD). Appendix B provides implementation and computational details (Table 3).

4.2 Control Performance and Learning Efficiency

Under driving-video backgrounds plus camera jitter, CaNI achieves higher mean return than every baseline on all five tasks (Figure 3). Its task-balanced mean is 715.5 versus 310.5 for R2-Dreamer. The gain is approximately 405 points (130.5%); CsDreamer scores 228.5. The matched backbone comparison isolates the training intervention's benefit with architecture and objectives fixed. Appendix D.2 reports per-task means and sample standard deviations.

On the driving-video benchmark without jitter, CaNI and R2-Dreamer have comparable 900k–990k final-window returns (853 versus 844), while CaNI learns faster (Figure 4(a)). Its 0–1M time-averaged return exceeds R2-Dreamer's on all five tasks, averaging 698.7 ± 11.2 versus 605.2 ± 60.0. Figure 5(a) shows the larger final-return gap with camera displacement. Appendices D.3 and D.4 give aggregation, five-method curves, and task-level scores.

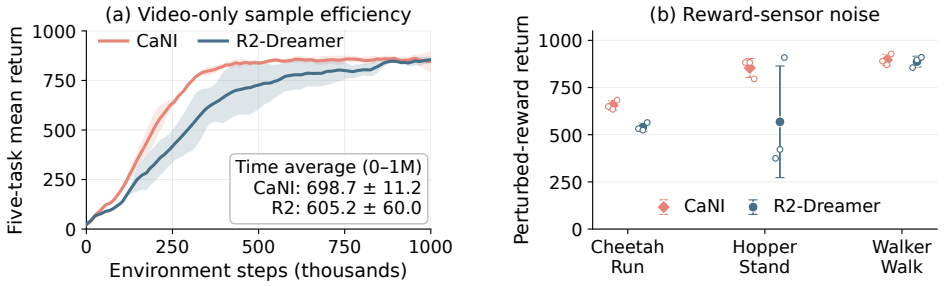

Figure 4: (a) Five-task mean video-only return. (b) Perturbed-reward returns. Bands and bars show three-seed sample SD; large and small markers show final-window means and replicates.

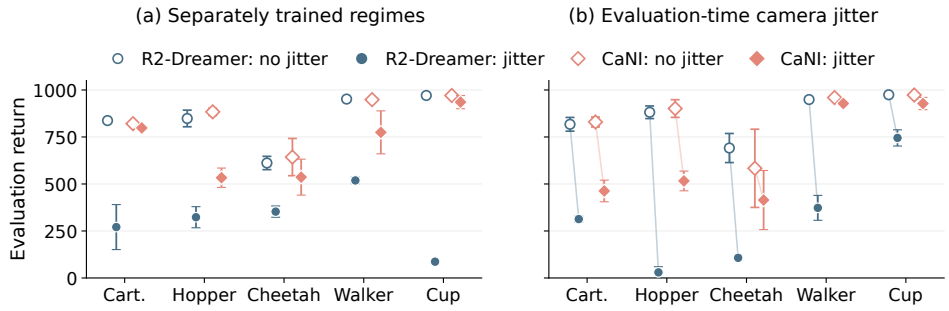

Figure 5: Effect of camera jitter: (a) separately trained agents, 900k–990k returns; (b) frozen 1M-step video-only policies. Lines in (b) connect evaluations with and without jitter for each task and method. Markers show three-seed means; error bars show sample SD.

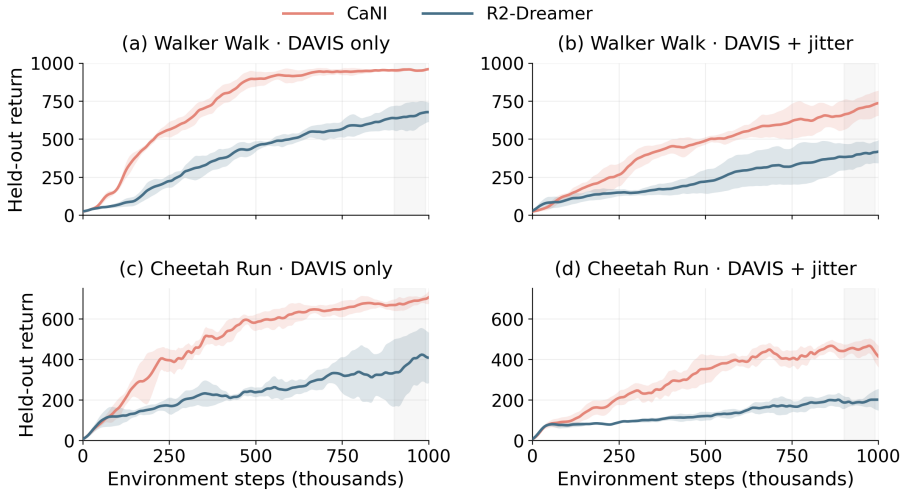

Figure 6: DAVIS transfer with and without jitter on disjoint evaluation videos. Curves show three-seed means; bands show sample SD. Gray shading marks 900k–990k steps.

As an empirical extension, we evaluate background-dependent corruption of the sensor used to compute reward. Training and evaluation both use the perturbed reward, which also defines the reported return. CaNI achieves higher mean returns than R2-Dreamer on Cheetah and Hopper, with near ties on Cartpole and Walker. Figure 4(b) summarizes Cheetah, Walker, and Hopper; Appendix D.5.2 gives all four task results and the protocol.

4.3 Frozen-policy robustness and background transfer

Evaluation-time camera shift. Introducing standard smooth jitter at evaluation changes the returns of fixed 1M-step policies trained without jitter from 863 to 314 for R2-Dreamer and from 849 to 650 for CaNI (Figure 5(b)). Jitter acts throughout complete rollouts, allowing recurrent states and actions to evolve; CaNI has higher return under evaluation-time jitter in all 15 matched task–seed comparisons.

Held-out background identities. With DAVIS backgrounds, we train on 60 sequences and evaluate on the disjoint 30-sequence validation split. With camera jitter, CaNI improves Walker from 398 to 697 (Figure 6(b)) and Cheetah from 194 to 456 (Figure 6(d)), extending the primary result to natural-video sequences not seen during training. The no-jitter comparisons in panels (a,c) also favor CaNI. Appendix D.5 gives the scores, and Appendix E visualizes the split.

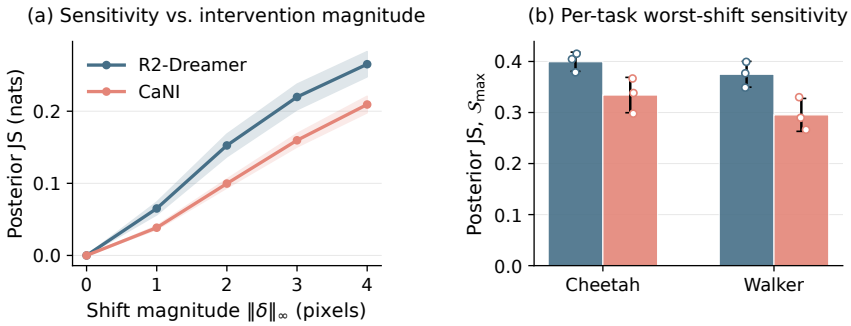

Figure 7: Posterior sensitivity on held-out DAVIS histories (JS in nats). (a) Cheetah and Walker are averaged equally within each replicate; shift magnitude is $\|\delta\|_\infty = \max(|\delta_x|, |\delta_y|)$ pixels. (b) Average statewise maximum sensitivity $\mathcal{S}_{\max}$. Means and sample SDs use three replicates; white markers show individual replicates.

4.4 Posterior intervention sensitivity

On shared held-out histories collected equally from both policies, each model fixes its own factual recurrent state and compares the factual posterior with posteriors induced by all 81 pad-4 translations. We average Jensen–Shannon (JS) divergence across categorical groups, compute the mean and maximum over translations at each state, and then average both statistics over histories. These group-averaged JS statistics summarize sensitivity over sampled histories and tested translations; Theorem 1 assumes a uniform posterior-TV bound across admissible states and intervention pairs. Appendix D.1 defines the statistics, displacement shells, and TV connection.

At fixed recurrent states, CaNI reduces the decoder-free backbone's mean sensitivity at every tested nonzero displacement magnitude (Figure 7). Relative reductions in mean sensitivity and average statewise maximum sensitivity are 22.9% and 16.3% on Cheetah, and 28.7% and 21.2% on Walker, respectively. The corresponding task-balanced reductions are 25.6% and 18.7%. This analysis complements the complete-rollout robustness tests in Section 4.3.

4.5 Intervention design and sensitivity

Temporal coupling. Figure 8 compares no augmentation, independent per-frame pad-4 shifts, and CaNI's sequence-shared shifts on Cheetah Run, Hopper Stand, and Walker Walk. Their task-balanced means are 394.1, 497.7, and 614.7. Sequence sharing has a higher mean on every task. The Cheetah difference is smaller and uncertain (536.3 ± 95.6 versus 502.8 ± 7.3), while the observed Hopper and Walker gains are 209.6 and 108.0 points. Appendix D.6 gives task-level scores.

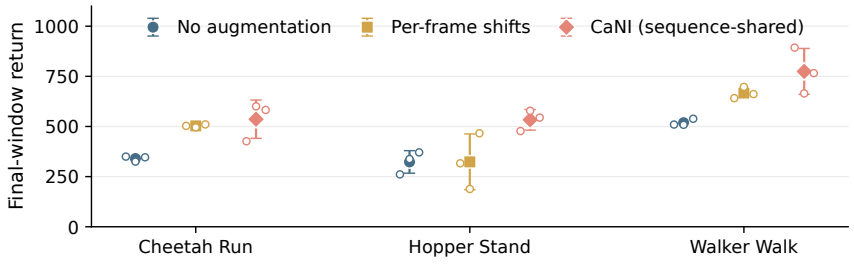

Figure 8: Temporal coupling with radius-four shifts. Large markers show final-window means; small markers show replicates. Error bars show three-seed sample SD.

Translation radius. We test sequence-shared translation radii of two, four, and eight pixels on Cheetah and Walker under the same nuisance setting. Mean final-window returns stay within 4.5% of the radius-four default on each task. Different radii attain the highest task means, suggesting limited sensitivity over the tested range. Appendix D.8 gives the curves and scores.

Intervention family. We compare sequence-shared translation and brightness/contrast under driving-video backgrounds plus camera jitter. Translation has higher mean returns on Cheetah, Walker, and Cup Catch (Table 1). Its advantage is largest on Cup (935.5 versus 335.1), where photometric scores vary substantially; the Walker gap is modest relative to across-seed variability. Brightness/contrast also improves over no augmentation on all three tasks. These results favor camera-aligned translation within the tested setting. Appendix D.7 gives augmentation settings and full curves.

Table 1: Sequence-shared augmentation: three-seed mean $\pm$ sample SD. R2-Dreamer control replicates differ from the primary comparison on Cheetah and Walker; see Appendix B.4.

Training variant	Cheetah Run	Walker Walk	Cup Catch
R2-Dreamer (family control)	340.3 ± 13.5	503.0 ± 39.2	86.6 ± 1.3
Brightness/contrast	407.7 ± 36.4	700.9 ± 111.4	335.1 ± 339.5
CaNI (radius 4)	536.3 ± 95.6	774.6 ± 114.4	935.5 ± 35.1

5 Related Work

Dreamer-style agents learn latent dynamics for imagined policy training (Hafner et al., 2025). DreamerPro and R2-Dreamer replace pixel reconstruction with representation objectives (Deng et al., 2022; Morihira et al., 2026), sharing the motivation of joint-embedding predictive architectures (JEPAs) (LeCun, 2022), including LeWorldModel and Causal-JEPA (Maes et al., 2026; Nam et al., 2026). Distractor-robust methods factor controllable or reward-relevant state from nuisance information (Wang et al., 2022; Fu et al., 2021; Liu et al., 2023). CaNI connects nuisance mechanisms to intervention design, recurrent-policy robustness, and matched-backbone training.

Image augmentation improves pixel-based reinforcement learning (RL) (Laskin et al., 2020b; Yarats et al., 2021; Hansen & Wang, 2021). DreamerPro applies sequence-consistent shifts through recurrent inference and initializes imagined policy learning from inferred states (Deng et al., 2022). CaNI uses the supplied nuisance mechanism to guide this training intervention. Causal analyses formalize augmentation through nuisance or style interventions (Mitrovic et al., 2021; von Kügelgen et al., 2021); here the preserved object is the joint controlled transition, reward, and continuation kernel, linking observation changes to sequential decisions.

Zhang et al. (2020, Theorem 5) bound return degradation through action-distribution TV under state-observation perturbations. Our result controls both actions and future memory through posterior stability under controlled-process-preserving channels. State-abstraction work links representations to reward and transition preservation (Gelada et al., 2019; Zhang et al., 2021); with exact posterior stability, control sufficiency yields this stronger abstraction. Appendix C gives extended comparisons with visual world models, augmentation, and causal representation learning.

6 Conclusion

CaNI shows how a specified nuisance mechanism can guide a simple replay intervention for recurrent world models. On an unchanged decoder-free backbone, sequence-consistent translation improves control under camera jitter, accelerates learning on the driving-video benchmark without jitter, and improves generalization to held-out DAVIS backgrounds. Frozen-policy evaluations and posterior diagnostics provide complementary evidence of robustness. The theory identifies uniform posterior stability as a sufficient condition for recurrent-policy return stability. Together, these findings highlight replay observation design as a useful complement to decoder-free representation objectives.

Limitations and future work. Bounded translation can discard control cues near image boundaries, such as the ball in Cup Catch; its suitability depends on their recoverability from transformed histories. Reward-sensor-noise benefits remain empirical, outside the observation-only theory. Exact abstraction additionally requires learned-state control sufficiency. Linking augmented optimization to uniform posterior stability and control sufficiency is a direction for future work.

AI USE STATEMENT

In this work, we used generative artificial intelligence (AI) tools to aid and polish writing and assist with literature retrieval and discovery. All AI-assisted content was reviewed and verified by the authors. The authors take full responsibility for the final content of the submission.

ETHICS STATEMENT

This work studies simulated visual control using established benchmark tasks and does not involve human participants or personal data. We are not aware of direct societal harms specific to the proposed method beyond the general limitations and potential misuse of reinforcement-learning systems.

REPRODUCIBILITY STATEMENT

Sections 2 and 3 specify the assumptions, theoretical results, intervention distribution, and training objective; Appendix A provides complete proofs. Appendix B documents baseline implementation sources, visual-environment construction, training configurations, evaluation protocols, and computational requirements. Appendix D details the posterior analysis, background splits, reward-sensor protocol, ablations, and result aggregation; Appendix E illustrates the sequence-shift construction. We will publicly release the CaNI source code and experiment configurations upon acceptance.

REFERENCES

Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019. URL https://arxiv.org/abs/1907.02893.

Simon Bing, Tom Hochsprung, Jonas Wahl, Urmi Ninad, and Jakob Runge. Invariance & causal representation learning: Prospects and limitations. *Transactions on Machine Learning Research*, 2024. URL https://openreview.net/forum?id=lpOC6s4BcM.

Joao Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the Kinetics dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6299–6308, 2017. URL https://openaccess.thecvf.com/content_cvpr_2017/html/Carreira_Quo_Vadis_Action_CVPR_2017_paper.html.

Xiaocong Chen, Siyu Wang, and Lina Yao. Continuity-regularized flow matching for offline reinforcement learning. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026a. URL https://icml.cc/virtual/2026/poster/61224.

Yuxin Chen, Jianglan Wei, Chenfeng Xu, Boyi Li, Masayoshi Tomizuka, Andrea Bajcsy, and Thomas Tian. Reimagination with test-time observation interventions: Distractor-robust world model predictions for visual model predictive control. In *2026 IEEE International Conference on Robotics and Automation*. IEEE, 2026b. URL https://ras.papercept.net/conferences/conferences/ICRA26/program/ICRA26_ContentListWeb_5.html.

Fei Deng, Ingook Jang, and Sungjin Ahn. DreamerPro: Reconstruction-free model-based reinforcement learning with prototypical representations. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 4956–4975. PMLR, 2022. URL https://proceedings.mlr.press/v162/deng22a.html.

Jiameng Fan and Wenchao Li. DRIBO: Robust deep reinforcement learning via multi-view information bottleneck. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 6074–6102. PMLR, 2022. URL https://proceedings.mlr.press/v162/fan22b.html.

Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov decision processes. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, pp. 162–169. AUAI Press, 2004. URL https://dblp.org/rec/conf/uai/FernsPP04.

Xiang Fu, Ge Yang, Pulkit Agrawal, and Tommi Jaakkola. Learning task informed abstractions. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3480–3491. PMLR, 2021. URL https://proceedings.mlr.press/v139/fu21b.html.

Carles Gelada, Saurabh Kumar, Jacob Buckman, Ofir Nachum, and Marc G. Bellemare. DeepMDP: Learning continuous latent space models for representation learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2170–2179. PMLR, 2019. URL https://proceedings.mlr.press/v97/gelada19a.html.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse control tasks through world models. *Nature*, 640:647–653, 2025. doi: 10.1038/s41586-025-08744-2. URL https://www.nature.com/articles/s41586-025-08744-2.

Nicklas Hansen and Xiaolong Wang. Generalization in reinforcement learning by soft data augmentation. In *2021 IEEE International Conference on Robotics and Automation*, pp. 13611–13617. IEEE, 2021. doi: 10.1109/ICRA48506.2021.9561103.

Nicklas Hansen, Hao Su, and Xiaolong Wang. Stabilizing deep Q-learning with ConvNets and vision transformers under data augmentation. In *Advances in Neural Information Processing Systems*, volume 34, pp. 3680–3693, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/1e0f65eb20acbfb27ee05ddc000b50ec-Abstract.html.

Miles Hutson, Isaac Kauvar, and Nick Haber. Policy-shaped prediction: Avoiding distractions in model-based reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 13124–13148, 2024. doi: 10.52202/079017-0418. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/17af43527227c5c96db0f8d4c6aadc4e-Abstract-Conference.html.

Michael Laskin, Aravind Srinivas, and Pieter Abbeel. CURL: Contrastive unsupervised representations for reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 5639–5650. PMLR, 2020a. URL https://proceedings.mlr.press/v119/laskin20a.html.

Misha Laskin, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, and Aravind Srinivas. Reinforcement learning with augmented data. In *Advances in Neural Information Processing Systems*, volume 33, pp. 19884–19895, 2020b. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/e615c82aba461681ade82da2da38004a-Abstract.html.

Yann LeCun. A path towards autonomous machine intelligence, 2022. URL https://openreview.net/forum?id=BZ5a1r-kVsf. Position paper, version 0.9.2.

Lihong Li, Thomas J. Walsh, and Michael L. Littman. Towards a unified theory of state abstraction for MDPs. In *Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics*, 2006. URL https://dblp.org/rec/conf/isaim/LiWL06.

Yuren Liu, Biwei Huang, Zhengmao Zhu, Honglong Tian, Mingming Gong, Yang Yu, and Kun Zhang. Learning world models with identifiable factorization. In *Advances in Neural Information Processing Systems*, volume 36, pp. 31831–31864, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/65496a4902252d301cdf219339bfbf9e-Abstract-Conference.html.

Lucas Maes, Quentin Le Lidec, Damien Scieur, Yann LeCun, and Randall Balestriero. LeWorldModel: Stable end-to-end joint-embedding predictive architecture from pixels. *arXiv preprint arXiv:2603.19312*, 2026. URL https://arxiv.org/abs/2603.19312.

Jovana Mitrovic, Brian McWilliams, Jacob Walker, Lars Buesing, and Charles Blundell. Representation learning via invariant causal mechanisms. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=9p2ekP904Rs.

Naoki Morihira, Amal Nahar, Kartik Bharadwaj, Yasuhiro Kato, Akinobu Hayashi, and Tatsuya Harada. R2-Dreamer: Redundancy-reduced world models without decoders or augmentation. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Je2QqXrcQq.

Heejeong Nam, Quentin Le Lidec, Lucas Maes, Yann LeCun, and Randall Balestriero. Causal-JEPA: Learning world models through object-level latent masking. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/63623.

Tung D Nguyen, Rui Shu, Tuan Pham, Hung Bui, and Stefano Ermon. Temporal predictive coding for model-based planning in latent space. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8130–8139. PMLR, 2021. URL https://proceedings.mlr.press/v139/nguyen21h.html.

Minting Pan, Xiangming Zhu, Yunbo Wang, and Xiaokang Yang. Iso-Dream: Isolating and leveraging noncontrollable visual dynamics in world models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23178–23191, 2022. doi: 10.52202/068431-1684. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/9316769afaaeeaad42a9e3633b14e801-Abstract-Conference.html.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009. doi: 10.1017/CBO9780511803161.

Silviu Pitis, Elliot Creager, and Animesh Garg. Counterfactual data augmentation using locally factored dynamics. In *Advances in Neural Information Processing Systems*, volume 33, pp. 3976–3990, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/294e09f267683c7ddc6cc5134a7e68a8-Abstract.html.

Jordi Pont-Tuset, Federico Perazzi, Sergi Caelles, Pablo Arbeláez, Alexander Sorkine-Hornung, and Luc Van Gool. The 2017 DAVIS challenge on video object segmentation. *arXiv preprint arXiv:1704.00675*, 2017. URL https://arxiv.org/abs/1704.00675.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954.

Austin Stone, Oscar Ramirez, Kurt Konolige, and Rico Jonschkowski. The distracting control suite—a challenging benchmark for reinforcement learning from pixels. *arXiv preprint arXiv:2101.02722*, 2021. URL https://arxiv.org/abs/2101.02722.

Yuval Tassa, Yotam Doron, Alistair Muldal, Tom Erez, Yazhe Li, Diego de Las Casas, David Budden, Abbas Abdolmaleki, Josh Merel, Andrew Lefrancq, Timothy Lillicrap, and Martin Riedmiller. DeepMind control suite. *arXiv preprint arXiv:1801.00690*, 2018. URL https://arxiv.org/abs/1801.00690.

Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 23–30. IEEE, 2017. doi: 10.1109/IROS.2017.8202133.

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello. Self-supervised learning with data augmentations provably isolates content from style. In *Advances in Neural Information Processing Systems*, volume 34, pp. 16451–16467, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/8929c70f8d710e412d38da624b21c3c8-Abstract.html.

Siyu Wang, Xiaocong Chen, and Lina Yao. On causally disentangled state representation learning for reinforcement learning based recommender systems. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 2390–2399. ACM, 2024a. doi: 10.1145/3627673.3679674. URL https://doi.org/10.1145/3627673.3679674.

Siyu Wang, Xiaocong Chen, Mingming Gong, Yong Li, Quan Z. Sheng, and Lina Yao. Reward-preserving counterfactual state editing for offline reinforcement learning. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/66163.

Tongzhou Wang, Simon Du, Antonio Torralba, Phillip Isola, Amy Zhang, and Yuandong Tian. Denoised MDPs: Learning world models better than the world itself. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 22591–22612. PMLR, 2022. URL https://proceedings.mlr.press/v162/wang22c.html.

Yucen Wang, Shenghua Wan, Le Gan, Shuai Feng, and De-Chuan Zhan. AD3: Implicit action is the key for world models to distinguish the diverse visual distractors. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 51546–51568. PMLR, 2024b. URL https://proceedings.mlr.press/v235/wang24bq.html.

Denis Yarats, Ilya Kostrikov, and Rob Fergus. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=GY6-6sTvGaf.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow Twins: Self-supervised learning via redundancy reduction. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12310–12320. PMLR, 2021. URL https://proceedings.mlr.press/v139/zbontar21a.html.

Amy Zhang, Rowan McAllister, Roberto Calandra, Yarin Gal, and Sergey Levine. Learning invariant representations for reinforcement learning without reconstruction. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=-2FCwDKRREu.

Huan Zhang, Hongge Chen, Chaowei Xiao, Bo Li, Mingyan Liu, Duane Boning, and Cho-Jui Hsieh. Robust deep reinforcement learning against adversarial perturbations on state observations. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21024–21037, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f0eb6568ea114ba6e293f903c34d7488-Abstract.html.

Shaowei Zhang, Jiahan Cao, Dian Cheng, Xunlan Zhou, Shenghua Wan, Le Gan, and De-Chuan Zhan. Leveraging conditional dependence for efficient world model denoising. In *Advances in Neural Information Processing Systems*, volume 38, pp. 76655–76688, 2025. doi: 10.52202/085713-2573. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/6f01b2a921c0a4dc34609613f4f74b8e-Abstract-Conference.html.

Chuning Zhu, Max Simchowitz, Siri Gadipudi, and Abhishek Gupta. RePo: Resilient model-based reinforcement learning by regularizing posterior predictability. In *Advances in Neural Information Processing Systems*, volume 36, pp. 32445–32467, 2023. doi: 10.52202/075280-1409. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/6692e1b0e8a31e8de84bd90ad4d8d9e0-Abstract-Conference.html.

A Proofs for Causal Nuisance Robustness and Exact Abstraction

Proof of Theorem 1. Couple the two processes from the same draw of the initial physical state and recurrent memory. While their physical states and recurrent memories agree, hold the nonintervened variables and exogenous randomness fixed as above, then maximally couple the representation kernels induced by the two selected channels. Equation 4 makes the conditional probability of unequal latent draws at most ε. When the latent draws agree, shared policy randomness gives the same action; coupling the common joint kernel K gives the same reward, continuation, and next physical state, and Assumption 1 gives the same next recurrent memory.

Let D_t be the event that the coupled latent draws have disagreed at or before time t, with $D_{-1} = \emptyset$. Superscript c denotes an event's complement, and $\Pr$ denotes probability under this coupling. Conditional on D_{t-1}^c, the probability of a first disagreement at time t is at most ε, hence

$$\Pr(D_t) \le 1 - (1-\varepsilon)^{t+1} \le (t+1)\varepsilon.$$

The rewards agree on D_t^c and each lies in $[0, R_{\max}]$, so

$$\begin{aligned} |J_Q(\pi) - J_{Q'}(\pi)| &\le \sum_{t=0}^{\infty} \gamma^t R_{\max} \Pr(D_t) \\ &\le R_{\max}\varepsilon \sum_{t=0}^{\infty} \gamma^t (t+1) = \frac{R_{\max}\varepsilon}{(1-\gamma)^2}. \end{aligned}$$

For the undiscounted return $J_Q^T(\pi) = \mathbb{E}_Q^\pi[\sum_{t=0}^{T-1} R_t]$ over T decisions, the same argument gives

$$\left|J_Q^T(\pi) - J_{Q'}^T(\pi)\right| \le R_{\max}\varepsilon \sum_{t=0}^{T-1} (t+1) = \frac{R_{\max}\varepsilon T(T+1)}{2}.$$

If $\varepsilon = 0$, $\Pr(D_t) = 0$ for every t. The coupling therefore gives identical physical states, learned states, actions, rewards, and continuations at every step, proving equality of their joint trajectory laws. □

Proof of Corollary 1. The full history $\mathcal{H}_t$ includes past observations, actions, rewards and continuation signals, the initial memory, and internal latent draws through time t; hence S_t is determined by $\mathcal{H}_t$. Fix a supported schedule Q and history-based policy μ. Define a possibly time-dependent abstract policy by

$$\overline{\mu}_t(\cdot \mid s) = \mathrm{Law}_Q^\mu(A_t \mid S_t = s).$$

The full-history sufficiency assumption and the tower property give the same conditional reward, continuation, and next-state kernel $\overline{P}$ under both policies. Starting from the same learned-state distribution, induction therefore gives identical time-indexed state–action marginals and expected returns. Thus optimizing over abstract policies loses no value relative to policies using the complete history. Conversely, sampling Z_t from Φ and applying an abstract policy to $S_t = (H_t, Z_t)$ implements its history-based lift.

The kernel $\overline{P}$ is common to all supported schedules. At $\varepsilon = 0$, Theorem 1 also couples their initial learned states and subsequent trajectories under any fixed abstract policy. Consequently, an optimal abstract policy lifts to an optimal history-based policy under each schedule, with the same value for the common initial physical-state law. □

B Experimental Setup and Implementation

B.1 Baseline methods

CaNI builds on the official R2-Dreamer codebase;[1] DreamerV3 and DreamerPro use the PyTorch implementations supplied in that repository. iFactor[2] and CsDreamer[3] use their authors' implementations with benchmark-specific task, visual-environment, and evaluation adapters; their learning settings are summarized below.

R2-Dreamer. R2-Dreamer (Morihira et al., 2026) replaces pixel reconstruction with a Barlow Twins (Zbontar et al., 2021) objective between image embeddings and projected recurrent states. The objective aligns corresponding feature dimensions and reduces cross-feature redundancy, providing representation learning without image augmentation.

[1] https://github.com/NM512/r2dreamer
[2] https://github.com/AlexLiuyuren/IFactor
[3] https://github.com/Zhang-Shaowei/CsDreamer

DreamerV3. DreamerV3 (Hafner et al., 2025) learns a categorical RSSM through observation reconstruction, reward prediction, and continuation prediction. Its actor and critic learn from trajectories imagined with the world model.

DreamerPro. DreamerPro (Deng et al., 2022) learns trainable prototypes and predicts their assignments from recurrent states and augmented observations instead of reconstructing pixels. Our configuration forms two views per replay sequence using sequence-shared shifts with a maximum displacement of three pixels and bilinear interpolation.

iFactor. iFactor (Liu et al., 2023) organizes the latent state into four blocks defined by action influence and reward relevance, allowing dependencies between blocks. Policy learning uses both controllable and uncontrollable reward-relevant components.

CsDreamer. CsDreamer (Zhang et al., 2025) uses a collider-structured RSSM to capture conditional dependence between task-relevant and task-irrelevant latent factors given observations. A decoupling regularizer limits information overlap between the factors during world-model training.

B.2 Visual environment and camera process

The primary benchmark extends the DMC visual-distractor environment of Denoised MDPs (Wang et al., 2022). Observations are 64×64 red–green–blue (RGB) images, with backgrounds drawn from 12 grayscale videos in the Kinetics-400 (Carreira & Zisserman, 2017) driving-car class and composited using the released foreground mask. Camera jitter translates the complete image along a temporally correlated path driven by random acceleration and damped velocity. The initial per-axis displacement standard deviation is 6.36 pixels; a restoring force acts beyond radius 9 pixels, without a hard displacement bound. The process advances twice between observations at action repeat two. Subpixel translations use bilinear interpolation and black borders.

B.3 Training configurations and CaNI objective

CaNI retains R2-Dreamer's architecture and complete world-model and behavior-learning objective, including its original loss weights. Table 2 compares training settings. Before encoding, each replay sequence is replicate-padded by four pixels and translated by one integer displacement sampled uniformly from $\{-4, \dots, 4\}^2$, shared across all timesteps. Actions, rewards, and continuation labels retain their recorded values; acting and evaluation use environment observations without additional CaNI augmentation. The matched R2-Dreamer control shares the architecture, objective, optimization, and evaluation protocol, with the training intervention disabled.

Translation geometry. Radius-four shifts retain the central 56×56 region of each 64×64 source frame, including objects wholly contained in that region. The training proxy uses sequence-shared nearest-neighbor integer shifts and replicate padding; the benchmark uses time-varying subpixel shifts and black borders.

Table 2: Training settings for the primary comparison. Batch denotes replay sequences; length and imagination horizon count transitions.

Method	Batch × length	Horizon	Optimizer	Learning rate model; actor/critic
CaNI, R2-Dreamer, DreamerV3, DreamerPro	16×64	15	LaProp	4×10^{-5} (joint)
iFactor	25×50	12	Adam	6×10^{-4}; 8×10^{-5}
CsDreamer	16×64	15	Adam	10^{-4}; 3×10^{-5}

Optimization and baseline settings. The common-codebase methods and CsDreamer use 512 replay transitions per collected transition. iFactor performs 100 minibatch updates every 1,000 collection steps. All use action repeat two. The CsDreamer Walker and Hopper aggregates each include one run resumed from saved model weights with reinitialized optimizer state.

Configuration selection. Each method uses a fixed configuration across the five primary tasks. CaNI inherits R2-Dreamer's learning hyperparameters; its final intervention design followed exploratory development on overlapping tasks. The temporal-sharing, augmentation-family, and radius studies examine the final method with the remaining settings fixed. Appendix D.7 specifies the brightness/contrast distribution and strength.

B.4 Evaluation and computational profile

Step counts and plotting. CaNI, R2-Dreamer, DreamerV3, DreamerPro, and CsDreamer count simulator steps: with action repeat two, 1M reported steps correspond to approximately 500k agent decisions. iFactor retains its implementation's decision-step counter, so 1M reported steps correspond to approximately 2M simulator steps. Its curves and final-window scores use this native counter.

Displayed learning curves use centered moving averages with replicated endpoints: nine points in Figures 1, 3, and 9, and five in Figure 12. Figures 3 and 9 smooth interpolated task means and SDs and the resulting task-aggregate curves. Final-window statistics use raw evaluations.

CaNI and R2-Dreamer sample categorical posterior latents during return evaluation, including frozen-policy tests, and select the actor distribution's mode conditional on the sampled learned state. They use independent evaluation random streams; iFactor's video-only and reward-sensor-noise comparisons follow its released training-seed evaluation convention. Learning curves report evaluation returns.

Primary and ablation scores average 900k–990k evaluations over three training replicates. The Cheetah temporal and augmentation-family ablations share R2-Dreamer controls that differ from the primary comparison; the Walker augmentation-family study also uses alternative controls. Other matching conditions reuse their reported aggregates.

Table 3: Computational profile under driving-video backgrounds plus camera jitter.

Method	Parameters (M)	Training time (h)	Reported steps/s
CaNI	11.840–11.844	21.4–37.8	7.8–10.0
R2-Dreamer	11.840–11.844	10.8–22.3	10.6–28.8
DreamerV3	10.490–10.494	13.2–30.8	10.9–23.2
DreamerPro	9.666–9.670	18.5–37.0	10.5–16.7
iFactor	9.177–9.184	41.6–48.4	5.7–6.7
CsDreamer	33.266–33.274	169.0–170.4	1.7–1.8

Parameter counts are trainable weights in millions; ranges reflect action dimensions. Throughput uses the native step counters defined above. Times and throughput are not hardware-controlled comparisons. CsDreamer timings cover Cheetah and Walker.

CaNI adds no trainable parameters; its translation cost scales linearly with the number of pixels in a replay batch.

C Extended Related Work

Visual-distraction benchmarks. The Distracting Control Suite extends DMC with controlled changes to backgrounds, colors, and rendered camera poses (Stone et al., 2021). Denoised MDPs instead studies DMC and RoboDesk variants while organizing observation information by controllability and reward relevance (Wang et al., 2022). The primary environment in this paper follows the Denoised MDP implementation lineage and adds the post-render smooth image-plane translation process specified in Appendix B.

World models for separating task-relevant and nuisance dynamics. TIA separates reward-correlated visual features from distractors through cooperative reconstruction with an adversarial reward objective (Fu et al., 2021). Denoised MDPs factor information by controllability and reward relevance, Iso-Dream separates controllable and noncontrollable dynamics using inverse-dynamics supervision, and iFactor models four latent categories defined by their interactions with actions and

rewards (Wang et al., 2022; Pan et al., 2022; Liu et al., 2023). RePo promotes representations predictive of dynamics and reward under an information-flow constraint and adds reward-free encoder alignment at test time (Zhu et al., 2023). AD3 infers implicit distractor actions and trains separated world models for heterogeneous and homogeneous distractors (Wang et al., 2024b). Policy-Shaped Prediction combines pretrained segmentation, task-aware reconstruction, and adversarial learning to focus model capacity on policy-relevant content (Hutson et al., 2024). CsDreamer models the conditional dependence between task-relevant and task-irrelevant latent components using a collider-structured RSSM and decoupling regularization (Zhang et al., 2025). CaNI uses a specified nuisance mechanism to guide observation-intervention design on the unchanged decoder-free backbone.

Decoder-free and temporal representation learning. Temporal predictive coding (TPC) learns recurrent representations for planning under natural-video distractions (Nguyen et al., 2021). DreamerV3 reconstructs observations, whereas DreamerPro and R2-Dreamer replace pixel reconstruction with prototype-based and redundancy-reduction objectives, respectively (Hafner et al., 2025; Deng et al., 2022; Morihira et al., 2026). DreamerPro forms two shifted views of each replay sequence, keeping each view's augmentation consistent across time (Deng et al., 2022). CURL learns representations by contrasting augmented observations (Laskin et al., 2020a); DRIBO extends multi-view information bottleneck learning to sequential observations through an auxiliary representation objective (Fan & Li, 2022).

This decoder-free motivation is shared by JEPAs: learning prediction targets in representation space allows models to omit unpredictable observation detail (LeCun, 2022). LeWorldModel jointly trains an encoder and predictive world model from pixels using next-embedding prediction and a Gaussian embedding regularizer (Maes et al., 2026). Causal-JEPA masks object-level latent histories to encourage prediction from other objects, introducing a causal inductive bias through controlled observability (Nam et al., 2026). CaNI addresses observation-channel robustness by using a supplied nuisance SCM to guide augmentation design while preserving the recorded actions, rewards, and continuation labels. Section 4.4 shows that the decoder-free R2-Dreamer backbone retains sensitivity to camera-shift proxies, motivating explicit training over the nuisance family.

Augmentation and observation intervention. RAD and DrQ establish the utility of simple crops and translations for pixel control; SODA decouples augmented representation learning from policy optimization, while SVEA stabilizes off-policy Q-learning under strong augmentation (Laskin et al., 2020b; Yarats et al., 2021; Hansen & Wang, 2021; Hansen et al., 2021). DreamerPro applies sequence-consistent shifts through recurrent inference and initializes imagined policy learning from inferred states (Deng et al., 2022). CaNI builds on these established ingredients with an SCM-based intervention formulation, a conditional recurrent-policy robustness analysis, and controlled evidence on the unchanged R2-Dreamer backbone. Reimagination with Observation Intervention detects and removes novel distractors at test time before re-compositing them into predicted videos (Chen et al., 2026b); CaNI intervenes during training under a specified nuisance family.

Causal foundations of augmentation. ReLIC interprets augmentations as interventions on style variables and regularizes proxy predictions for invariance across views (Mitrovic et al., 2021). von Kügelgen et al. (2021) give conditions under which augmented views identify invariant content up to an invertible mapping, allowing causal dependencies between content and style. In RL, CoDA uses locally independent causal mechanisms to recombine replay transitions into valid counterfactual experiences (Pitis et al., 2020), while Wang et al. (2026) study reward-preserving counterfactual state editing for offline RL. CaNI specializes causal augmentation to observation channels that preserve the joint controlled kernel, connecting posterior stability to recurrent-policy return and training over transformed observations of a fixed physical trajectory.

Robustness, abstraction, and causal invariance. For generative offline policies, Chen et al. (2026a) use a continuity-equation-based regularizer to promote smoothness and stability along action-generation flows. Zhang et al. (2020, Theorem 5) bound return degradation through action-distribution TV under state-observation perturbations of a Markov policy. Our recurrent formulation applies this divergence-to-return principle to posterior kernels: at a common memory state, the shared policy and recurrent update bound the TV of the joint action and next memory by posterior TV. Coupling this joint evolution controls return changes across causal observation-channel schedules.

Classical state-abstraction and bisimulation results characterize when reward and transition preservation support planning in a reduced state space, and quantitative metrics connect deviations from bisimulation to value differences (Li et al., 2006; Ferns et al., 2004). DeepMDP and deep bisimulation extend these ideas to learned continuous representations (Gelada et al., 2019; Zhang et al., 2021). CIDS uses conditional mutual information to identify action-influenced state components and their causal ancestors for recommendation policy learning (Wang et al., 2024a). Invariant risk minimization seeks representations admitting a common optimal predictor across training environments (Arjovsky et al., 2019). Causal representation learning motivates invariance across mechanism changes, while also showing that invariance alone need not identify latent causal variables (Pearl, 2009; Schölkopf et al., 2021; Bing et al., 2024). Corollary 1 combines exact nuisance invariance with the control-sufficiency condition for an exact abstraction and optimal-control preservation.

D Additional Experimental Results

D.1 Posterior intervention-sensitivity results

We evaluate both models on a shared bank of ten 500-step histories per task and training replicate, contributed equally by their frozen 1M-step policies on DAVIS validation backgrounds. Histories use posterior modes and deterministic actions; each model reconstructs its recurrent states using posterior modes and scores 1,000 states. Return evaluations sample posterior latents (Appendix B.4). We average within each replicate before reporting across-replicate means and sample SDs. Task-balanced percentage reductions are computed relative to R2-Dreamer after averaging sensitivities equally across tasks and replicates, using full precision before rounding.

For each shared held-out history, model M, with representation parameters θ, reconstructs and holds fixed its own factual recurrent state h_t^M while the current observation is branched through the tested shift set $\mathcal{U} = \{-4, \dots, 4\}^2$. Each branch is evaluated from that factual state. For categorical group $g \in \{1, \dots, 32\}$, let $q_{t,g}^0$ and $q_{t,g}^\delta$ be its probability vectors under $\Phi_\theta(\cdot \mid h_t^M, o_t)$ and $\Phi_\theta(\cdot \mid h_t^M, T_\delta(o_t))$, after the model's configured uniform-probability mixing (unimix). Superscript 0 denotes the factual observation. The full factorized posteriors are $q_t^0 = \prod_{g=1}^{32} q_{t,g}^0$ and $q_t^\delta = \prod_{g=1}^{32} q_{t,g}^\delta$; all quantities below are computed separately for each model and seed.

For categorical probability vectors p, q, define the Kullback–Leibler divergence $D_{\mathrm{KL}}(p\|q) = \sum_k p_k \log(p_k/q_k)$ and $D_{\mathrm{JS}}(p, q) = \frac{1}{2} D_{\mathrm{KL}}(p\|\overline{q}) + \frac{1}{2} D_{\mathrm{KL}}(q\|\overline{q})$, where k indexes their classes and $\overline{q} = (p + q)/2$. Logarithms are natural, so divergences are measured in nats. The group-averaged Jensen–Shannon divergence is

$$d_t(\delta) = \frac{1}{32} \sum_{g=1}^{32} D_{\mathrm{JS}}\big(q_{t,g}^0, q_{t,g}^\delta\big) . \tag{9}$$

Zero sensitivity across all groups is equivalent to equality of the factorized kernels. With $\mathbb{E}_t$ denoting the empirical average over the 1,000 scored states within one model and seed, we summarize mean and worst-shift sensitivity as

$$\mathcal{S}_{\mathrm{mean}} = \mathbb{E}_t \left[\frac{1}{81} \sum_{\delta \in \mathcal{U}} d_t(\delta) \right], \qquad \mathcal{S}_{\max} = \mathbb{E}_t \left[\max_{\delta \in \mathcal{U}} d_t(\delta) \right] . \tag{10}$$

The maximum is taken over translations separately at each state before averaging over histories. Figure 7(a) groups shifts by $m(\delta) = \|\delta\|_\infty$ and averages $d_t(\delta)$ over every direction in each displacement shell.

Table 4 summarizes both sensitivity measures.

The groupwise statistic has a distributional connection to Theorem 1. For any factorized posteriors $q = \prod_{g=1}^{32} q_g$ and $q' = \prod_{g=1}^{32} q'_g$, with categorical group distributions q_g, q'_g, Jensen–Shannon divergence with natural logarithms and Pinsker's inequality give

$$\mathrm{TV}(q, q') \le \min \left\{ 1, \sqrt{2 \sum_{g=1}^{32} D_{\mathrm{JS}}(q_g, q'_g)} \right\} . \tag{11}$$

Indeed, joint JS divergence is at most the sum of the groupwise divergences, and $\mathrm{TV}^2 \leq 2D_{\mathrm{JS}}$. For a factual-to-shift comparison, Equation 11 becomes $\mathrm{TV}(q_t^0, q_t^\delta) \leq \min\{1, \sqrt{64 d_t(\delta)}\}$; two shifted channels can be compared through the factual channel and the triangle inequality for total variation. At the observed JS magnitudes this upper bound evaluates to one. The reported averages quantify comparative sensitivity on sampled histories; Theorem 1 assumes a uniform bound ε over all supported pairs.

Table 4: Posterior intervention sensitivity for Cheetah Run and Walker Walk trained with DAVIS backgrounds plus camera jitter and evaluated on held-out DAVIS validation backgrounds. Values are means $\pm$ sample standard deviations across three training seeds, in nats. Lower is better.

Task	Method	$\mathcal{S}_{\mathrm{mean}}$	$\mathcal{S}_{\mathrm{max}}$
Cheetah Run	R2-Dreamer	$.2162 \pm .0138$	$.3995 \pm .0188$
	CaNI	$.1666 \pm .0211$	$.3342 \pm .0346$
Walker Walk	R2-Dreamer	$.1967 \pm .0203$	$.3750 \pm .0253$
	CaNI	$.1403 \pm .0180$	$.2953 \pm .0322$

CaNI reduces both sensitivity measures on both tasks, consistent with the intended invariance to the camera-aligned intervention family.

D.2 Aggregate primary results

Table 5: Primary benchmark returns under driving-video backgrounds plus camera jitter. Scores average 900k–990k evaluations (mean $\pm$ sample standard deviation, three seeds). Control sets: Appendix B.4.

Method	Cartpole Swingup	Hopper Stand	Cheetah Run	Walker Walk	Cup Catch
CaNI	798.1 ± 5.2	533.2 ± 51.5	536.3 ± 95.6	774.6 ± 114.4	935.5 ± 35.1
R2-Dreamer	270.7 ± 119.7	323.2 ± 56.2	353.0 ± 30.0	518.8 ± 17.0	86.6 ± 1.3
DreamerV3	154.8 ± 13.7	12.4 ± 6.9	229.4 ± 37.7	422.0 ± 14.7	153.8 ± 56.1
DreamerPro	179.1 ± 11.5	3.5 ± 2.1	169.3 ± 41.0	356.9 ± 99.5	120.6 ± 46.5
iFactor	132.3 ± 25.4	18.1 ± 19.3	200.6 ± 53.9	183.9 ± 43.9	117.6 ± 21.4
CsDreamer	160.7 ± 16.5	91.1 ± 78.7	347.8 ± 204.8	481.4 ± 118.3	61.6 ± 38.3

CaNI leads on every task. Its task-balanced mean is 715.5 versus 310.5 for R2-Dreamer, the strongest baseline, and 228.5 for CsDreamer: gains of 130.5% and 213.1%, respectively. CsDreamer's across-seed variability is largest on Cheetah.

D.3 Full video-only comparison

Figure 9 reports five-method learning curves under driving-video backgrounds without camera jitter. R2-Dreamer and CaNI reach comparable final-window returns, while CaNI achieves higher return over training on every task. We quantify sample efficiency for these two matched methods as the trapezoidal time average of evaluation return over 0–1M environment steps, computed per run and then averaged equally across three seeds and five tasks. The task-balanced values are 698.7 ± 11.2 for CaNI and 605.2 ± 60.0 for R2-Dreamer (mean $\pm$ sample standard deviation across the three seed-wise task averages); the corresponding task means are 696 versus 670 on Cartpole, 696 versus 487 on Hopper, 522 versus 425 on Cheetah, 786 versus 693 on Walker, and 793 versus 751 on Cup.

For final-window return, DreamerV3 scores 596.2 ± 64.4 on Walker and 163.4 ± 25.2 on Cup; DreamerPro scores 540.6 ± 63.0 and 188.6 ± 12.6, respectively. On Cheetah, DreamerPro scores 426.1 ± 60.2 and iFactor scores 150.0 ± 74.5; iFactor scores 5.0 ± 1.2 on Hopper. These values are means $\pm$ sample standard deviations across three training seeds.

D.4 Camera-jitter attribution and evaluation shifts

Figure 5 compares scores with and without camera jitter while holding the driving-video protocol fixed. In the video-only setting, final-window returns averaged over 900k–990k steps for R2-

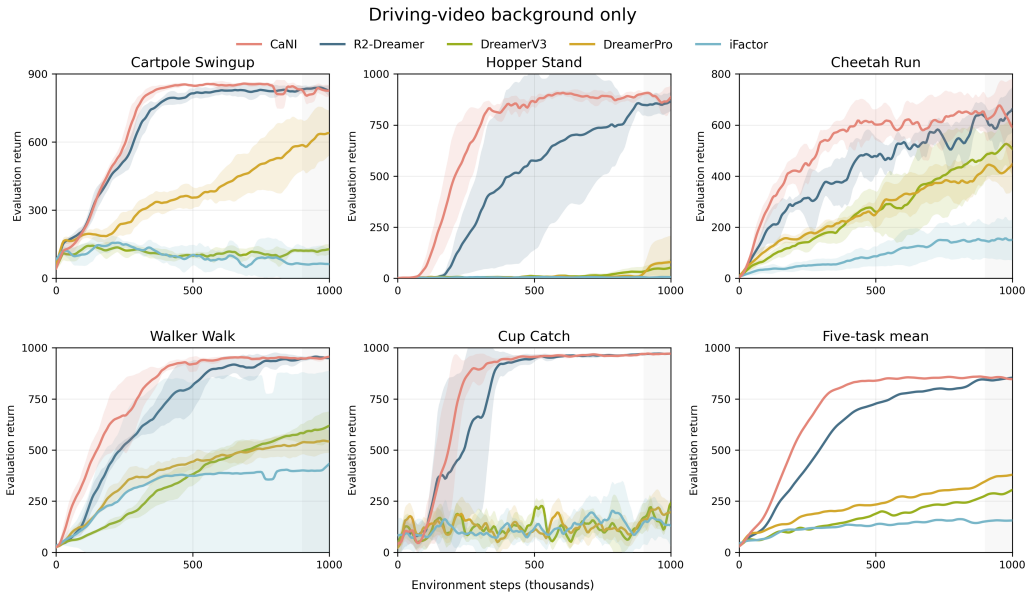

Figure 9: Cartpole Swingup, Hopper Stand, Cheetah Run, Walker Walk, and Cup Catch under driving-video backgrounds without camera jitter. Curves show means across three training seeds; bands indicate sample standard deviations. The final panel averages task means equally for all five methods.

Dreamer versus CaNI are 837 versus 821 on Cartpole, 849 versus 884 on Hopper, 612 versus 643 on Cheetah, 952 versus 949 on Walker, and 971 versus 970 on Cup.

For the evaluation-time shift in Figure 5(b), exact-1M video-only versus video-plus-jitter returns are 817 versus 313 on Cartpole, 881 versus 30 on Hopper, 691 versus 107 on Cheetah, 949 versus 373 on Walker, and 975 versus 745 on Cup for R2-Dreamer. The corresponding CaNI returns are 830 versus 463, 901 versus 516, 583 versus 415, 960 versus 928, and 973 versus 928. Each value averages three training seeds; both evaluation conditions use ten evaluation episodes per checkpoint.

D.5 DAVIS GENERALIZATION AND REWARD-SENSOR NOISE

The DAVIS experiments use training videos and held-out validation videos from disjoint splits. Figure 6 organizes Walker Walk and Cheetah Run by task and camera-jitter regime.

D.5.1 DAVIS BACKGROUND REGIMES

Without camera jitter, Walker final-window returns are 954.1 ± 7.9 for CaNI and 653.9 ± 84.9 for R2-Dreamer, reported as mean $\pm$ sample standard deviation over three seeds. Cheetah returns are 683.7 ± 17.3 for CaNI versus 383.0 ± 144.7 for R2-Dreamer, each over three seeds. With camera jitter, the corresponding three-seed means are 697 versus 398 on Walker and 456 versus 194 on Cheetah. CaNI has higher observed returns under both DAVIS background regimes on both tasks.

D.5.2 REWARD-SENSOR-NOISE RESULTS

The sensor-noise protocol combines video backgrounds with corrupted reward measurements and disables camera jitter. Let $m \in [0, 255]$ be the mean intensity of a fixed 16×16 background-video patch. A clean scalar sensor reading s_{clean} is replaced by $s_{\text{meas}} = s_{\text{clean}} + c_{\text{task}}(m/255 - 1/2)$ when computing reward. The sensor and scale c_{task} are cart position and 2 for Cartpole, forward speed and 10 for Cheetah, torso height and 1.2 for Walker, and torso height and 0.6 for Hopper. The background determines the perturbation at each physics step. Both training rewards and reported evaluation returns use this perturbed reward measurement.

Table 6 summarizes reward-sensor-noise returns on Cheetah, Walker, and Hopper. CaNI has a higher observed mean than R2-Dreamer on Cheetah and Hopper, while Walker is nearly tied. DreamerV3, DreamerPro, and iFactor have lower observed means than both on all three tasks. Hopper shows substantial variability for R2-Dreamer, DreamerPro, and iFactor. On Cartpole Swingup, R2-Dreamer and CaNI are approximately tied at 810.0 ± 11.6 and 805.0 ± 16.5, while iFactor scores 54.9 ± 90.5 (mean $\pm$ sample standard deviation, $n = 3$).

Table 6: Final-window returns with driving-video backgrounds plus reward-sensor noise. Values are mean $\pm$ sample standard deviation over three seeds, averaging each seed's evaluations at 900k–990k reported steps.

Method	Cheetah Run	Walker Walk	Hopper Stand
CaNI	655.7 ± 25.2	896.2 ± 29.6	853.7 ± 50.0
R2-Dreamer	540.3 ± 21.5	887.4 ± 28.2	568.5 ± 295.8
DreamerV3	398.5 ± 31.0	592.0 ± 94.6	48.4 ± 60.2
DreamerPro	273.8 ± 17.4	335.6 ± 34.8	133.8 ± 220.4
iFactor	284.8 ± 211.4	451.5 ± 75.3	338.7 ± 295.4

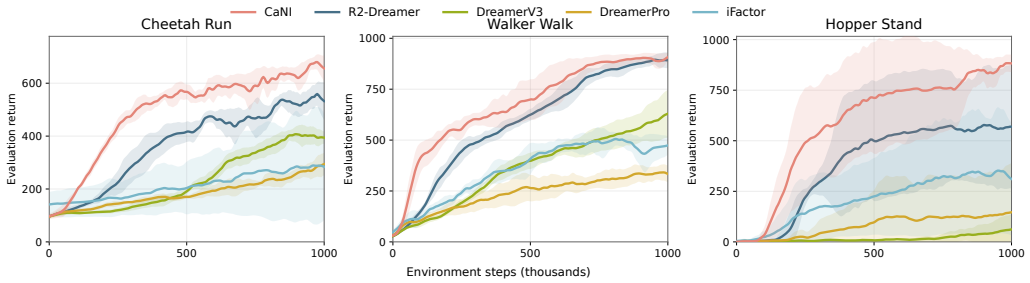

Figure 10: Performance under driving-video backgrounds plus reward-sensor noise. Curves show means and bands indicate sample standard deviations across three training seeds.

D.6 Temporal-coupling ablation

The temporal-coupling ablation uses Cheetah Run, Hopper Stand, and Walker Walk under driving-video backgrounds plus camera jitter. It compares no augmentation, independent per-frame random shifts, and sequence-shared random shifts (CaNI). The two shift conditions use the same pad-4 translation family and differ only in whether the sampled transform is shared across time.

Table 7: Temporal-coupling ablation under driving-video backgrounds plus camera jitter. Scores average 900k–990k evaluations (mean $\pm$ sample standard deviation, three seeds). R2-Dreamer control replicates differ from the primary comparison on Cheetah; see Appendix B.4.

Task	Condition	Score
Cheetah Run	R2-Dreamer (no augmentation)	340.3 ± 13.5
	Per-frame shift	502.8 ± 7.3
	Sequence-shared shift (CaNI)	536.3 ± 95.6
Hopper Stand	R2-Dreamer (no augmentation)	323.2 ± 56.2
	Per-frame shift	323.6 ± 139.0
	Sequence-shared shift (CaNI)	533.2 ± 51.5
Walker Walk	R2-Dreamer (no augmentation)	518.8 ± 17.0
	Per-frame shift	666.6 ± 28.4
	Sequence-shared shift (CaNI)	774.6 ± 114.4

The task-balanced means for no augmentation, per-frame shifts, and sequence-shared shifts are 394.1, 497.7, and 614.7, respectively. Sequence-shared augmentation has a higher observed mean

than per-frame augmentation on all three tasks, with an average gain of 117.0 points; the Cheetah advantage is smaller and has greater uncertainty.

D.7 Augmentation-family comparison

For image intensities normalized to $[0, 1]$, the photometric alternative samples a brightness offset uniformly from $[-0.10, 0.10]$ and a contrast multiplier from $[0.80, 1.20]$, independently for each replay sequence. Contrast acts about intensity 0.5, followed by brightness addition and clipping to $[0, 1]$. Both parameters are shared across frames and color channels. Translation is disabled in this condition; the backbone and training objective are retained.

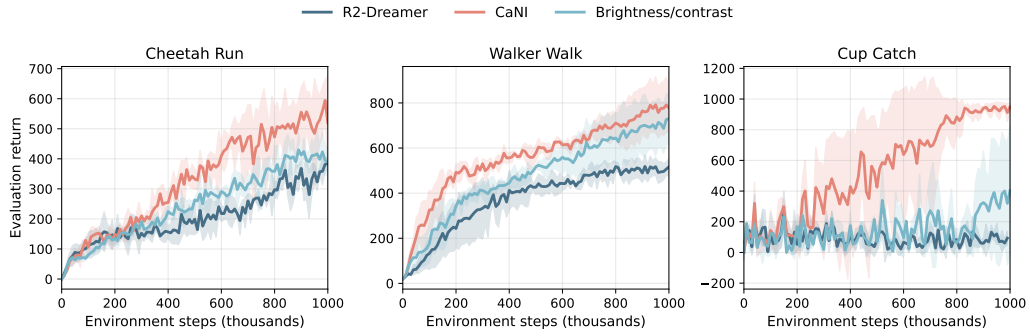

Figure 11: Augmentation-family comparison under driving-video backgrounds plus camera jitter. Curves and bands show means and sample standard deviations over three seeds. CaNI uses radius-four translation; both augmentation families are sequence-shared.

CaNI has the highest mean final-window return on all three tasks (Table 1). Its advantage over brightness/contrast is largest on Cup Catch, where photometric augmentation has substantial variability, and smallest on Walker. Brightness/contrast improves mean return over no augmentation on every task.

D.8 Translation-radius sensitivity

Figure 12 shows the learning curves for the radius study in Section 4.5; Table 8 reports their final-window returns. All three radii use three training seeds on each task, with each sampled translation shared across the replay sequence. The primary CaNI configuration uses radius four.

Table 8: Translation radius under video backgrounds and jitter (three-seed mean ± sample SD).

CaNI radius	Cheetah Run	Walker Walk
2 pixels	512.9 ± 29.9	807.7 ± 129.3
4 pixels (default)	536.3 ± 95.6	774.6 ± 114.4
8 pixels	540.7 ± 61.3	740.0 ± 97.3

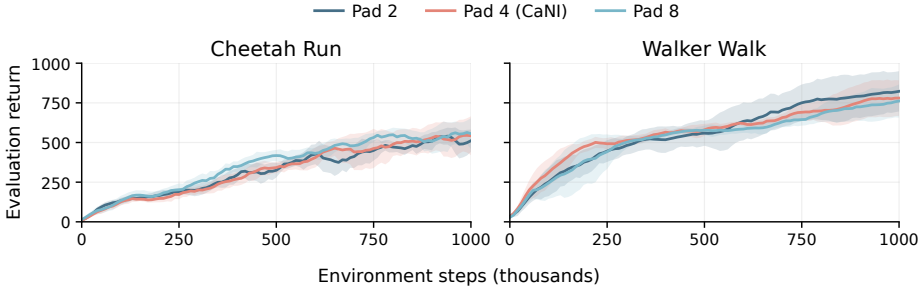

Figure 12: Translation-radius sensitivity: smoothed three-seed means with sample SD bands.

E QUALITATIVE PROTOCOL FIGURES

Figure 13 uses fixed evaluation episodes. For each task it shows the observed frame and samples from the replicate-padded translation family used during training. The displayed views share the same physical state, background, and control labels. Source observations are 64×64 pixels and are enlarged for inspection.

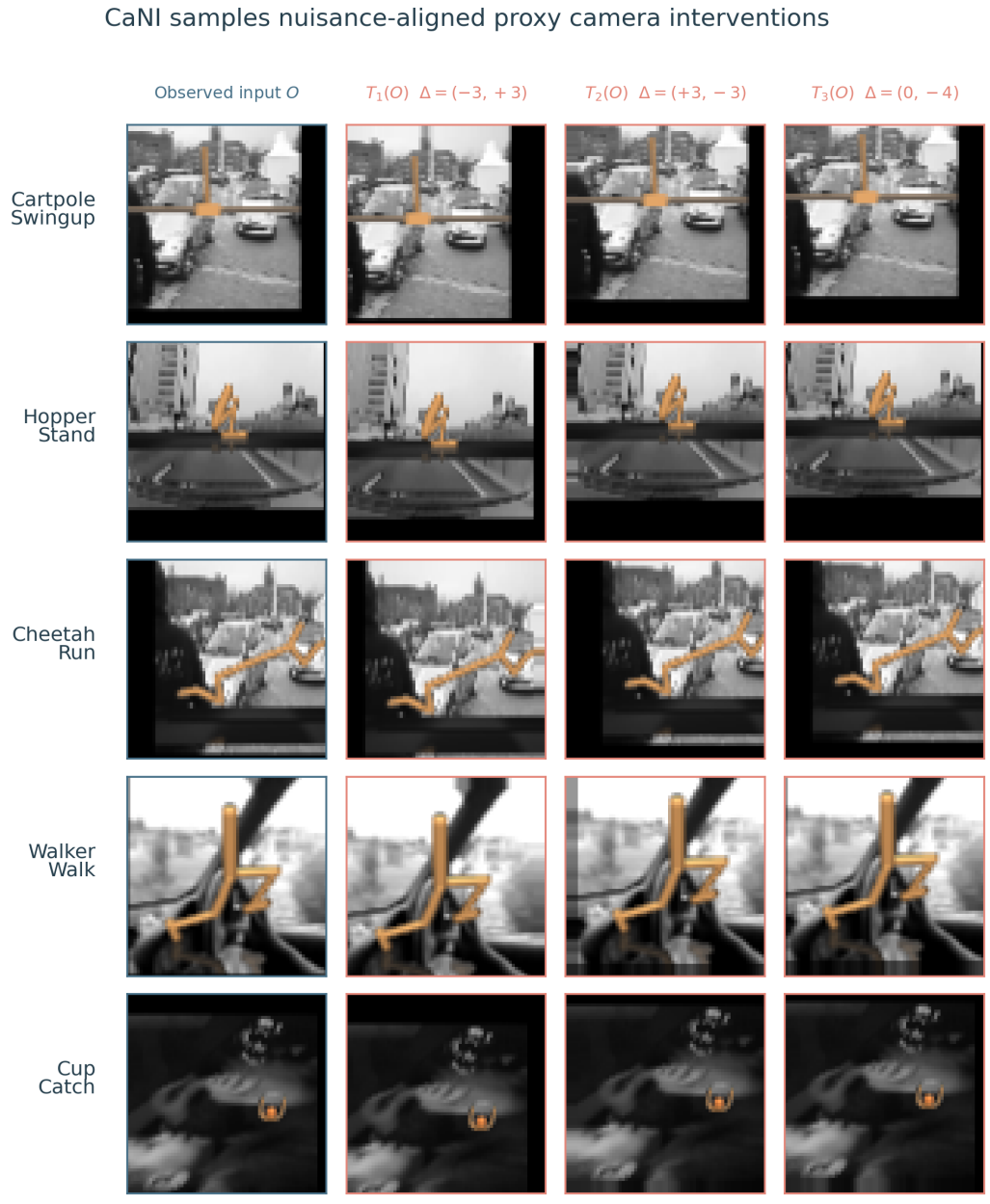

Figure 13: Qualitative CaNI interventions for the five primary tasks under driving-video backgrounds plus camera jitter. Each row shows a replay observation O and three indexed translations T_1, T_2, T_3 from the family T_δ. Labels $\Delta = -\delta$ give visible horizontal/vertical displacement in pixels (right/down positive), opposite to the sampling offset. Physical state, action, reward, and continuation labels are unchanged.

Figure 14 isolates the same intervention on held-out DAVIS backgrounds with benchmark camera jitter. It uses CaNI evaluations on Cheetah Run and Walker Walk and exposes the transformation and its replicate-padding boundary behavior directly.



Figure 14: CaNI proxy interventions for Cheetah Run and Walker Walk on held-out DAVIS validation backgrounds with camera jitter. The left panel is the observed policy input; the remaining panels apply replicate-padded translations with visible pixel displacements $\Delta = -\delta$ (right/down positive). Within each row, physical state, action, reward, and continuation labels are fixed.

Figure 15 illustrates the background generalization protocol with source frames from the environment's training and validation video splits.

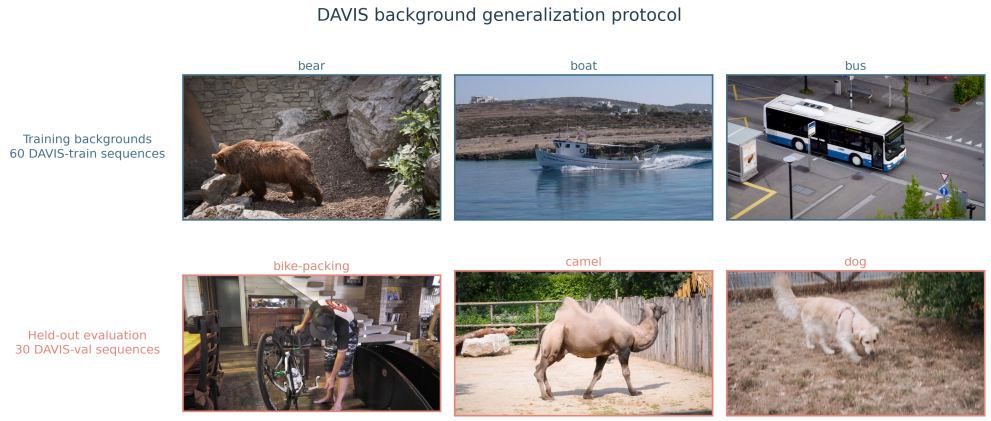

Figure 15: Representative DAVIS background frames from the disjoint protocol partitions. Collection environments draw only from the 60 training sequences (top), while evaluation environments draw only from the 30 held-out validation sequences (bottom). Sequence names are shown above each example.

Figure 16 illustrates the driving-video protocol: collection and evaluation draw independent random streams from the same 12-video pool. The panel separates dynamic background composition from camera jitter by holding the Cheetah state and source frame fixed before applying a controlled image-plane translation.

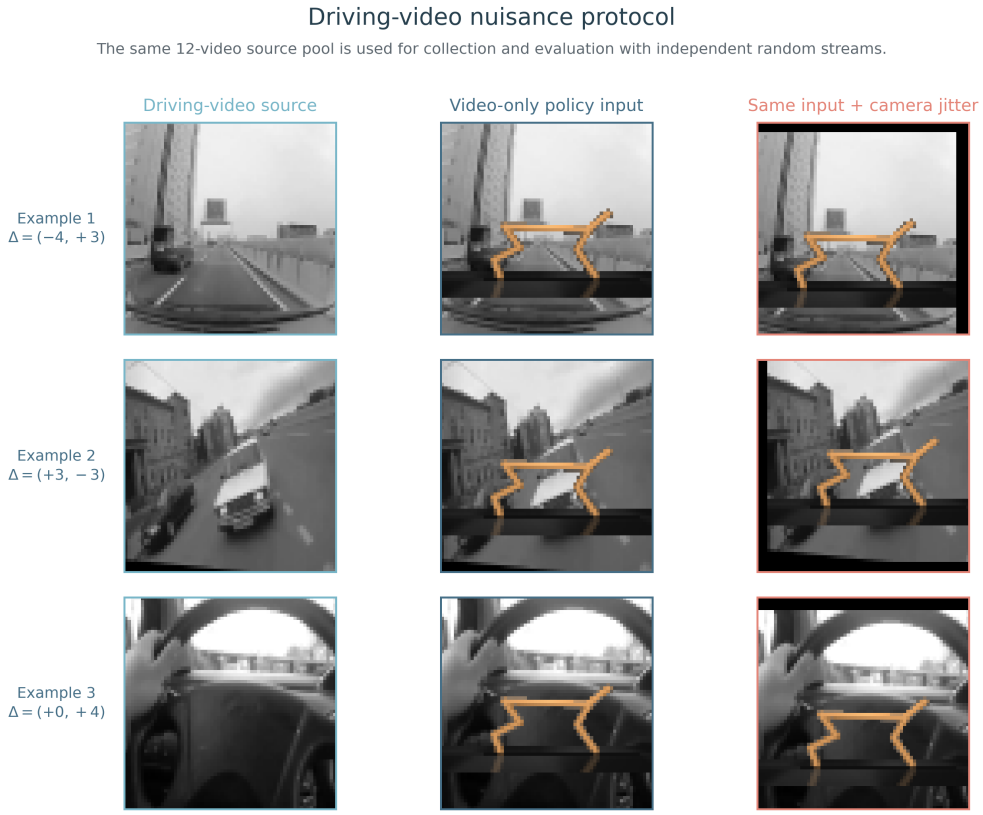

Figure 16: Driving-video nuisance construction. Each row fixes one Cheetah physics state and one grayscale source frame; the center column shows the video-only composite, and the right column applies a representative image-plane translation Δ in horizontal/vertical pixels (right/down positive) to illustrate camera jitter. Collection and evaluation use independent random streams from the same 12-video pool.