ANAMNESIS: TRAINING-FREE EPISODIC RECOLLECTION FOR AUTOREGRESSIVE WORLD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Autoregressive (AR) world models generate interactive worlds by conditioning on the recent past, yet they systematically forget it: scenes revisited after minutes of exploration are regenerated as if seen for the first time, and spatial memory collapses within the span of a single attention window. We argue that this failure has a precise phenomenological signature: current world models implement *retention* (the KV cache) and *protention* (action conditioning), but possess no *recollection*: the deliberate reactivation of earlier experience when the present re-encounters it. We introduce Anamnesis, a training-free episodic recollection mechanism for frozen AR world models: a buffer stores denoised latent states tagged with exact camera poses; a pose-gated detector recognizes revisits; and recognized revisits are re-rendered by re-noising the stored latent to a low level and granting the model a single refinement step. No gradients are updated. A KV-cache alternative (writing the anchor into the recency slot) leaves the output unchanged: recollection must enter through the generative channel, not the retention channel. To measure recollection we propose a closed-loop revisit benchmark with analytically exact pose ground truth from the model's own action interface, and an Excess Revisit Consistency metric that controls for the slow-motion shortcut. On the open-source LingBot-World model (one 4-GPU node), the baseline's excess is negligible (mean ERC-PSNR $+0.35$ dB; *negative* on the pan loop) while Anamnesis lifts it to $+2.24$ dB on all four loops ($p < 10^{-6}$), improving PSNR, LPIPS, and SSIM throughout, with a random-anchor control showing the effect is carried by pose-gated detection. Widening the KV window up to $6\times$ does not restore what the excursion erased: recollection, not longer retention, is the missing capability.

1 INTRODUCTION

A world model that cannot remember is not simulating a world; it is improvising one. Recent interactive video world models—Genie 3 (DeepMind Team, 2025), Oasis (Decart and Etched, 2024), Matrix-Game (He et al., 2025a), HY-WorldPlay (Tu et al., 2025), and LingBot-World (Robbyant Team, 2026)—generate minutes of responsive, controllable gameplay-like video. But their creators consistently report the same pathology: memory fades within roughly a minute. Genie 3 describes a "visual memory" of about one minute beyond which revisited regions emerge from "fog" (DeepMind Team, 2025); the Oasis team notes that with a 30-second training window the model "barely pays attention to frames beyond the past few"; LingBot-World reports emergent spatial memory at the one-minute scale as its qualitative frontier. For worlds meant to be *lived in* (explored, returned to, built upon), this is the central deficit.

We give this deficit a name borrowed from phenomenology. Husserl's analysis of time-consciousness (Husserl, 1964), recently mapped onto computational models of perception by Albarracin et al. (2022), distinguishes three structurally different ways a mind relates to its past: *retention*, the passive holding of the just-elapsed moment; *protention*, the anticipatory orientation toward the imminent; and *recollection*, the deliberate reactivation of an earlier experience when the present re-encounters it. Current AR world models implement the first two mechanically: the KV cache *is* retention, and action conditioning *is* protention. But nothing in the architecture re-collects: when the camera returns to a place seen a minute ago, the model has no operation that says *this is that*.

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.

The past is either still in the window (retention) or gone entirely. What is missing is precisely recollection, absent because the architecture has no episodic store and no retrieval trigger.

This gap is not merely engineering. The identity of a place across visits is the Ship of Theseus question made measurable: if the model regenerates the revisited scene from its generative prior rather than its own earlier experience, the "same" place is a different object wearing the same name. Loop closure—returning to a visited location and finding it unchanged—is the operational test of place identity, and it is exactly the test that existing world models fail (DeepMind Team, 2025; Wang et al., 2026b).

Why not train memory in? Existing remedies bake memory into the weights: WorldMem (Xu et al., 2025) adds memory-attention modules; ReWorld (Anonymous, 2026) trains pose-indexed landmark banks; DreamX-World (Wang et al., 2026a) trains a memory-conditioned persistence module; Matrix-Game 3 (He et al., 2026a) trains camera-aware retrieval. These are strong results, but they inherit three costs: training compute that few groups have; a memory budget fixed at training time; and, critically, *they cannot be applied to the frozen, released checkpoints that define the current open ecosystem*. A mechanism that grants recollection *post hoc* to frozen world models is therefore both a scientific probe (isolating recollection from all other learned capabilities) and a practical upgrade path. Concurrent training-free efforts stay on the retention side of this divide: they keep a compressed KV cache addressable (Wu et al., 2026), manage its growth (Zhao et al., 2026), or lock cached keys for editing (Zeng et al., 2026). None returns a *specific past episode* through the generative channel; §6.4 shows why that channel is the one that matters.

Contributions. This paper makes four contributions:

1. **A phenomenological diagnosis** of memory in AR world models: we formalize the retention/protention/recollection trichotomy, show that existing architectures implement exactly the first two, and connect the recollection gap to measurable loop-closure failure (§4).
2. **Anamnesis**, a training-free episodic recollection mechanism for frozen AR world models: a pose-indexed episodic buffer of denoised latent states, a pose-gated revisit detector with analytic ground truth, and two recollection injection mechanisms—*recollection by renoising* (SDEdit-style regeneration from the episodic anchor under the revisited pose) and *recollection by context override* (writing the anchor into the KV cache recency slot) (§4).
3. **A closed-loop revisit evaluation protocol** with analytically exact pose ground truth, constructed from the world model's own action string interface: an Excess Revisit Consistency (ERC) metric—which pairs every revisit measurement with a novel-view control at matched temporal distance, neutralizing the slow-motion shortcut identified by Du et al. (2026)—and a *memory-horizon* measurement that reduces persistence to a single curve over excursion length (§5). Relative to the contemporaneous MemoryGain calibration of R2M-Bench, our deltas are analytic pose ground truth, the horizon scalar, and method evaluation rather than model ranking.
4. **An empirical study** on the open-source LingBot-World checkpoint: baseline revisit quality is poor and its excess over a novel-view control negligible; Anamnesis lifts that excess six-fold ($p < 10^{-6}$) on a single 4×RTX-4090 node with zero training (anchored chunks execute fewer denoising steps, so recollection is cheaper than synthesis), and an attention-horizon sweep shows widening the window up to $6\times$ does not restore what the excursion erased, evidence that recollection, not longer retention, is the missing primitive (§6).

2 Related Work

Interactive video world models. GameNGen (Valevski et al., 2024) and Oasis (Decart and Etched, 2024) established neural game engines; Genie (Bruce et al., 2024) and Genie 3 (DeepMind Team, 2025) scaled promptable interactive worlds; The Matrix (Feng et al., 2024), HunyuanGameCraft (Li et al., 2025), Matrix-Game 1–3.5 (He et al., 2025a;b; 2026a;b), HY-WorldPlay (Tu et al., 2025), Yume (Mao et al., 2025), and LingBot-World (Robbyant Team, 2026) pushed duration, resolution, and action spaces. LingBot-World is, to our knowledge, the strongest fully open-weight AR world model: it fine-tunes a Wan2.2 14B MoE video DiT with Plücker-ray and WASD action conditioning, and adapts the bidirectional model to causal chunk-autoregressive rollout with diffusion forcing and sliding-window KV attention. We build on its released fast checkpoint.

Memory mechanisms in world models (trained). WorldMem (Xu et al., 2025) attends from current frames to a memory bank of past states; ReWorld (Anonymous, 2026) maintains a pose-indexed landmark bank with bounded per-head context windows and evaluates revisits with palindrome trajectories; DreamX-World (Wang et al., 2026a) conditions on retrieved memory states for scene persistence; Matrix-Game 3 (He et al., 2026a) retrieves memory patches by camera pose and reinjects residuals; Matrix-Game 3.5 (He et al., 2026b) disentangles static and dynamic content; HY-WorldPlay (Tu et al., 2025) compresses context into a reconstituted memory; Alaya-EVOKE (Hu et al., 2026) externalizes a world-state bank; Memory Forcing (Huang et al., 2025) trains a geometry-indexed spatial memory with point-to-frame retrieval; VRAG (Chen et al., 2025) retrains retrieval-augmented conditioning after showing that *naive* extended context and clean retrieved frames are largely ignored by video models; the training-side evidence points at conditioning-noise level as the binding variable, the knob §6.4 turns into an inference-time law in the frozen regime. These are *trained* mechanisms: the memory read path, its schedule, and its budget are learned or architected at training time, so none can be applied to a frozen released checkpoint. Anamnesis is complementary: it demonstrates that a substantial fraction of recollection is achievable *without any training*, by exploiting the denoising process itself as a memory decoder.

Training-free cache and context manipulation (concurrent). Concurrent to this work, three lines attack memory without retraining, all inside the *retention* channel: WorldTrace (Wu et al., 2026) makes a compressed KV cache addressable again by assigning in-distribution virtual RoPE positions to summary slots (with LoopBench for episodic recall), DensityKV (Zhao et al., 2026) manages the historical KV bank by token-level redundancy, and KV-Lock (Zeng et al., 2026) fuses cached background KVs into DiT self-attention for video editing. In LLMs, overriding a field's own KV entries is known to be largely ignored because prefill has already written the conclusion downstream (Li, 2026). Our KV-recency override (Mechanism B) is a world-model instance of this pattern: we find the same null result, and we show that recollection succeeds only when the past re-enters through the *generative* channel (Mechanism A, SDEdit-style re-noising of the episodic anchor), a route none of the cache-side methods takes. Closest in mechanism, RAGDP (Odonchimed et al., 2025) re-noises retrieved expert actions to accelerate diffusion policies, but for speed, not memory, with no revisit framing; Yang et al. (2024) calibrate SDEdit noise levels for video enhancement, not recollection.

Revisit benchmarks and the evaluation gap. MIND (Wang et al., 2026b) scores memory consistency on closed-loop revisited routes; WorldMark (Shi et al., 2026) measures local, global, and revisit memory at three timescales; R2M-Bench (Du et al., 2026) introduces revisit trajectories over 100 real scenes and documents the *slow-motion shortcut*: absolute revisit similarity is confounded by degenerate static generation, and proposes a Normalized Memory Ratio against novel-view baselines. Echo- Memory (Park et al., 2026) provides a controlled memory-capacity study. WBench (Fan et al., 2026) and WorldScore (Deng et al., 2025) audit world simulation broadly; SANA-WM-Bench (Research et al., 2026) defines revisit consistency as pixel metrics on geometrically thresholded revisit frame pairs ($< 0.5\,\text{m}$, $< 20^\circ$); Memory Forcing (Huang et al., 2025) evaluates trained spatial memory in Minecraft; WorldTrace's LoopBench (Wu et al., 2026) probes whether a *compressed* cache can reconstruct a detoured scene, a retention-capacity question, whereas we ask what a *frozen* model does when the pose itself returns. Our benchmark adopts the R2M-Bench lesson (novel-view controls) but differs in three respects: revisit poses are *analytically exact* (derived by construction from the action string interface, not estimated); the control is computed on the same rollout, yielding an excess metric that isolates recollection within a single generation; and the target of evaluation is a *method* under a frozen checkpoint, not a model ranking.

Episodic memory in RL and agents. Episodic control has a long history outside generative models: MFEC (Blundell et al., 2016) and NEC (Pritzel et al., 2017) store experience in differentiable neural dictionaries queried by k-NN; Never Give Up (Badia et al., 2020) uses k-NN novelty as an intrinsic signal; Go-Explore (Ecoffet et al., 2021) returns to archived states to continue exploration. These systems retrieve *values* or *novelty signals*; none had to regenerate pixels. Anamnesis transfers the archive-and-retrieve discipline to latent video generation, where the retrieved item must be fused into a diffusion rollout.

Philosophy of memory in machine systems. Phenomenological analyses of time-consciousness (Husserl, 1964) have been mapped onto active-inference architectures (Ramstead et al., 2022); object permanence and scene-persistence have been studied as binding problems in vision (Greff et al., 2020; Veerapaneni et al., 2019). The Ship of Theseus has been invoked constructively in machine learning, most directly to justify replacing the components of a trained network piece by piece while preserving its behavior (Subramaniam et al., 2025). No prior work uses the retention/protention/recollection trichotomy to diagnose—and mechanically repair—memory in video world models.

3 PRELIMINARIES

Chunk-autoregressive world models. Let $z_{1:T}$ be latent video frames and $a_{1:T}$ actions. A chunk-autoregressive world model generates a chunk of C latent frames at each step:

$$z^{(c)} \;=\; \mathrm{Denoise}_\theta\Big(\epsilon^{(c)};\, \mathrm{ctx}\big(z^{(<c)}\big),\, a^{(c)}\Big), \tag{1}$$

where $\mathrm{ctx}(\cdot)$ is a bounded attention context, in LingBot-World-Fast a sliding-window KV cache of the last W latent frames (window $W = L_{\mathrm{att}} \times n_{\mathrm{tok}}$ tokens for a local attention size of L_{att} latent frames at n_{tok} tokens per frame), possibly with sink tokens (Xiao et al., 2024). After denoising, the clean chunk $z_0^{(c)}$ is written back into the KV cache by a zero-noise "context forward" pass, so the window always contains the most recent W tokens of *generated* latent content.

Recollection is structurally absent. Two properties follow. First, the context is *purely temporal*: it contains the most recent frames, regardless of where the agent is. Second, the context is *lossy by eviction*: content older than the window is discarded, not archived. When the agent revisits a location visited more than W frames ago, the denoiser conditions on the recent past (elsewhere) plus its generative prior; the earlier visit contributes nothing. We call this the *recollection gap*.

Action strings and analytic poses. LingBot-World conditions generation on camera actions expressed as a compact string DSL—e.g. `"none-9,w-40,s-40,none-12"`: hold for 9 frames, move forward 40, back 40, hold 12—which the released code compiles deterministically into a per-frame camera trajectory $\{P_t \in SE(3)\}$ (WASD translation at fixed speed, IJKL rotation at fixed rate). Because the mapping is deterministic, *palindromic out-and-back strings return the camera to its departure pose with $\sim 10^{-16}$ translation error*: the ground-truth revisit frame index is known exactly, without any pose estimation. This is the foundation of our benchmark (§5).

4 ANAMNESIS: TRAINING-FREE EPISODIC RECOLLECTION

Anamnesis is a runtime wrapper around a frozen chunk-AR world model. It adds three components: an episodic buffer, a revisit detector, and two injection mechanisms. No parameters are trained; the wrapper runs on the released checkpoint as-is.

4.1 EPISODIC BUFFER

During rollout, each denoised chunk $z_0^{(c)}$ is already computed by the pipeline (it is written to the KV cache); Anamnesis stores it on CPU as well, tagged with the exact camera pose at the chunk's first latent frame, $p^{(c)} = P_{3c}$ (the compiled trajectory sampled at the latent-frame rate, one pose $P \in SE(3)$ per latent frame):

$$\mathcal{B} \;=\; \big\{\big(c,\; z_0^{(c)} \in \mathbb{R}^{16\times h\times w},\; p^{(c)} \in SE(3)\big)\big\}_{c=0}^{T/C-1}. \tag{2}$$

Storage is one bf16 latent chunk (≈0.5 MB for 480p) per step: negligible relative to the model, and, unlike pixel frames, the stored item is exactly the quantity the denoiser consumes.

4.2 POSE-GATED REVISIT DETECTION

At each chunk c, the detector compares the current pose $p^{(c)}$ against buffer entries, excluding recent chunks (a *recency margin* g chunks, so the departure context is not mistaken for a revisit):

$$s(c', c) = \max\Big(\frac{\|t_{c'} - t_c\|}{\tau_t},\ \frac{d_{\angle}(R_{c'}, R_c)}{\tau_r}\Big), \qquad \hat{c}(c) = \arg\min_{c' \le c-g} s(c', c), \tag{3}$$

and a revisit is declared iff $s(\hat{c}(c), c) \le 1$. Poses are exact (given), so detection is deterministic and threshold-interpretable: τ_t meters and τ_r degrees define what counts as "the same place." This is k-NN episodic detection (Pritzel et al., 2017; Badia et al., 2020) with analytic state, applied to world-model rollout.

4.3 RECOLLECTION INJECTION

Given a detected anchor $(\hat{c}, z_0^{(\hat{c})})$, the revisit chunk c must be generated *from* the recollected content rather than from pure noise. We instantiate two mechanisms.

Mechanism A: recollection by re-noising. Replace the revisit chunk's initial noise with the anchor, re-noised to an intermediate flow-matching timestep k^*:

$$z_{k^*}^{(c)} \ = \ \mathrm{AddNoise}\big(z_0^{(\hat{c})},\ k^*\big), \qquad \text{denoise from step } k^* \text{ under } (a^{(c)}, P^{(c)}). \tag{4}$$

The anchor supplies the scene layout; the final denoising steps under the current pose and action re-render it at the revisited viewpoint. This is SDEdit (Meng et al., 2022) applied *across time within one rollout*: the episodic anchor plays the role SDEdit assigns to a reference image. Setting k^* at the low-noise end of the schedule (our default: $k^*{=}t_3$, the last and least-noisy of the 4 inference steps) yields a single denoising step: recollection costs one forward pass per revisit.

Mechanism B: recollection by context override. The zero-noise context forward normally writes the chunk's own clean latent into the KV recency slot. When a revisit is detected for chunk $c{+}1$, Anamnesis instead writes the *anchor* latent $z_0^{(\hat{c})}$ into that slot at the end of chunk c:

$$\mathrm{KV}[c] \ \leftarrow \ z_0^{(\hat{c})} \quad \text{so that} \quad \mathrm{Attn}\big(z_{\text{noise}}^{(c+1)}\big) \ \ni \ \big\{\dots,\ z_0^{(\hat{c})}\big\}. \tag{5}$$

The revisited chunk then attends to recollected content as if it were the immediately preceding context. This exploits a property unique to AR world models: the KV cache is a *writable* working-memory slot, and its content determines what "retention" contains.

Controls. A *random* condition applies both mechanisms with a randomly chosen buffer entry (matched cost, no pose gating), isolating the contribution of retrieval from that of the injection machinery. The `none` condition is the frozen model unmodified.

4.4 WHAT ANAMNESIS DOES AND DOES NOT CLAIM

Anamnesis is a probe and a patch, not a solution to world memory. It claims: (i) recollection can be added to a frozen AR world model with zero training and measurable effect; and (ii) the phenomenological trichotomy is operationally useful: each term maps to a mechanism (retention = KV cache; protention = action conditioning; recollection = buffer + detection + injection). It does not handle dynamic-scene memory, long-horizon consistency beyond the anchor's fidelity, or trained architectures at scale (§8).

5 THE CLOSED-LOOP REVISIT BENCHMARK

Design. The world model's own action interface *is* the pose ground truth: every trajectory is an action string whose compiled pose sequence is exact by construction. We evaluate four canonical loops (§3): `outback_40` (out 40 frames, return, 93 frames total), `outback_80` (177 frames), `pan_45` (45° pan left, pan back, 105 frames), and `square_40` (a square loop returning to start, 177 frames). For each, the departure frame (last on-pose frame before excursion), the exact revisit frame, and the *novel frame* (maximum pose distance from departure) are computed from the compiled poses.

Metrics. Let $\mathrm{sim}_\triangle(f_1, f_2)$ be a similarity between frames. For departure frame d, revisit r, and novel frame n:

$$ERC_\triangle \;=\; \mathrm{sim}_\triangle(f_d, f_r) - \mathrm{sim}_\triangle(f_d, f_n), \qquad NMR_\triangle \;=\; \frac{\mathrm{sim}_\triangle(f_d, f_r)}{\mathrm{sim}_\triangle(f_d, f_n)}, \tag{6}$$

for $\triangle \in \{\mathrm{PSNR}, \mathrm{SSIM}, \mathrm{LPIPS}, \mathrm{DINO}\}$ (LPIPS differences are sign-flipped so that higher is better throughout). The novel-view control follows R2M-Bench's anti-shortcut lesson (Du et al., 2026): a model that degenerates to static video gains no ERC excess over its own drift to novel views; a true recollector scores positive ERC. A motion proxy (mean inter-frame difference) confirms comparable dynamics: pooled over the 32 scene–seed cells, Anamnesis rollouts move *more* than baseline (paired ratio 1.13; 1.03–1.20 per loop), so consistency gains are not bought with frozen frames.

Relation to concurrent benchmarks. Positioning against the revisit benchmarks above is given in §2. What they leave open is a *method*-level answer under a frozen checkpoint: exact pose ground truth, a same-rollout excess, and a horizon scalar. Those are precisely the tools §6.3 uses to falsify the capacity hypothesis and §6.4 to separate the two recollection channels.

Protocol. For each scene and trajectory, all conditions share the initial frame, prompt, seed, and attention window; only the recollection condition varies, and metrics are computed offline from saved videos. The main comparison covers 4 scenes $\times$ 4 loops $\times$ 2 conditions $\times$ 2 seeds (64 matched rollouts) plus random-anchor and KV controls; ablations add excursion-length (up to 333 frames), anchor-noise (3 levels), gate-jitter (2), and window-size (5) sweeps.

6 Experiments

Setup. We use the released LingBot-World-Fast checkpoint (nominal 14B-parameter DiT, bf16; the released fp32 serialization is ≈69 GiB, see appendix) with 4-step flow-matching inference, chunk size 3 latent frames, 480p (464×832), on a single node with 4×RTX-4090 (24 GB) using FSDP sharding (the fp32 checkpoint is cast to bf16 before sharding; see appendix). The local attention window is 3 latent frames (the minimum that fits 24 GB at this resolution); a horizon sweep at reduced resolution varies the window (§6.3). Detection defaults: τ_t=0.15 m, τ_r=10°, margin g=2 chunks. The novel-view control frame is selected by rotation-aware SE(3) distance (translation alone degenerates for rotation-only loops).

6.1 Baseline: the recollection gap is real

Table 1 ("none" rows) quantifies the gap on the frozen model, pooled over four scenes and two seeds: revisit PSNR is poor and barely excursion-dependent (11.97, 11.75, 11.95, 12.29 dB on the four loops), and ERC-PSNR is small overall (mean $+0.35$ dB; two-sided $p = 0.022$ against zero): positive but weak on outback-40/outback-80/square-40 ($+0.59$, $+0.40$, $+0.65$) and *negative* on pan-45 (-0.26): on the pan the model is less faithful to a remembered place than to a novel one. Its mean frame difference (0.04–0.06) rules out degenerate slow-motion (the R2M-Bench confound). Baseline revisit quality also degrades with excursion duration in LPIPS ($0.646 \rightarrow 0.678$), matching the qualitative "fog" reported for frontier systems.

6.2 Main results

Table 1 shows three findings. **(1) Recollection works.** Re-noising improves revisit PSNR on all four loops ($+2.61$, $+1.95$, $+1.67$, $+1.36$ dB; mean $+1.90$ dB over 32 scene–seed pairs) and improves LPIPS on all four. On outback-40, pooled SSIM rises from 0.319 to 0.467; the model regenerates the remembered scene rather than a plausible substitute. **(2) The detector, not the anchor budget, carries the effect.** The random control fires as many anchors or more (13 vs. 6 on outback-80) yet matches baseline pooled: mean ERC-PSNR $+0.10$, statistically indistinguishable both from zero ($p = 0.98$) and from the frozen baseline ($p = 0.43$), vs. $+0.35$ (none) and $+2.24$ (renoise); renoise beats random paired ($p < 10^{-3}$); on average, recollected frames beat the *novel*-view control, while randomly re-anchored frames do not. Part of the baseline's small positive excess is prior,

Loop	Cond	PSNR	LPIPS	ERC-PSNR	ERC-LPIPS	n
outback_40	none	11.97 ± 0.81	0.646 ± 0.009	0.59 ± 0.91	0.008 ± 0.013	(8)
outback_40	random	14.07 ± 1.96	0.547 ± 0.095	0.49 ± 1.94	-0.110 ± 0.102	(8)
outback_40	renoise	14.59 ± 1.28	0.465 ± 0.024	3.20 ± 1.52	0.188 ± 0.023	(8)
outback_40	kv	11.97 ± 0.82	0.646 ± 0.009	0.58 ± 0.93	0.008 ± 0.015	(8)
outback_80	none	11.75 ± 0.96	0.678 ± 0.017	0.40 ± 0.62	-0.007 ± 0.011	(8)
outback_80	random	13.34 ± 1.76	0.579 ± 0.067	0.24 ± 2.09	0.018 ± 0.066	(8)
outback_80	renoise	13.69 ± 1.56	0.593 ± 0.021	2.32 ± 2.24	0.077 ± 0.030	(8)
outback_80	kv	11.74 ± 0.95	0.677 ± 0.015	0.37 ± 0.62	-0.007 ± 0.011	(8)
pan_45	none	11.95 ± 0.70	0.622 ± 0.040	-0.26 ± 1.01	0.024 ± 0.019	(8)
pan_45	random	12.33 ± 1.16	0.585 ± 0.044	-0.67 ± 0.45	-0.012 ± 0.009	(8)
pan_45	renoise	13.62 ± 0.89	0.504 ± 0.031	1.42 ± 0.89	0.141 ± 0.015	(8)
pan_45	kv	11.92 ± 0.71	0.620 ± 0.041	-0.29 ± 1.00	0.026 ± 0.018	(8)
square_40	none	12.29 ± 0.81	0.666 ± 0.013	0.65 ± 0.36	0.016 ± 0.021	(8)
square_40	random	13.39 ± 1.77	0.578 ± 0.068	0.32 ± 2.09	0.020 ± 0.064	(8)
square_40	renoise	13.65 ± 1.52	0.592 ± 0.025	2.01 ± 2.02	0.089 ± 0.030	(8)
square_40	kv	12.32 ± 0.80	0.666 ± 0.013	0.68 ± 0.30	0.016 ± 0.021	(8)

Table 1: Revisit consistency (480p, window w=3), *mean ± std over 4 scenes (00,01,02,05) × 2 seeds*; n in parentheses. ERC = excess revisit consistency vs. the novel-view control (Eq. 6). Revisit pairs are pose-exact (translation error $\sim 10^{-16}$ m).

not memory: on loops that end where they began (out-back, square), any two renders of the start scene share layout by typicality, while the pan loop—whose revisit pose shares nothing with its excursion—exposes the pure recollection demand, and there the baseline is *negative*. The random control's large variance (± 2 dB) reflects the same heterogeneity: an accidental corridor hit helps, a distant anchor does nothing. **(3) KV recency override alone does nothing.** Rewriting the most recent context slot with an anchor latent leaves metrics indistinguishable from baseline. Under diffusion forcing, a single clean cache slot does not out-shout four tokens of fresh sampling noise; the SDEdit-style in-situ path is the operative mechanism. We report this null as evidence that the recollection channel matters, not the particular injection bus. **Multi-seed statistics.** Pooled over the 4-scene × 2-seed grid (paired by scene, loop, seed; one-sided Wilcoxon signed-rank), recollection beats the frozen baseline on ERC-PSNR (p¡0.001, $n = 32$) and ERC-LPIPS (p¡0.001, $n = 32$); the excess on ERC-DINO (p=0.069, $n = 32$) does not reach one-sided significance. Against the random-anchor control the advantage holds on ERC-PSNR (p¡0.001, $n = 32$) and ERC-LPIPS (p¡0.001, $n = 32$)—the pose gate, not anchor volume, carries the effect.

Three honest caveats. First, the SSIM *decrease* we initially observed for renoise on outback-80 and square-40 at scene 00, seed 0 ($0.266 \rightarrow 0.230$) did not survive pooling: across the 4-scene × 2-seed grid mean SSIM *rises* on all four loops ($+0.09$ to $+0.15$). A single cell can manufacture a consistency–fidelity trade that the full design does not support. Second, the semantic excess is mixed: ERC-DINO is positive for renoise on the loops where one anchor chunk covers most of the revisit content (pan-45 $+0.32$, outback-40 $+0.22$) but negative on outback-80 (-0.08) and square-40 (-0.12), where the revisit spans more chunks than a single re-noised anchor can restore: layout and palette reinstate faster than instance appearance (pooled $p = 0.069$). Third, on outback-40 corridor geometry places every anchor near the current camera, so the gate is weak there; we flag it as *corridor revisit* and treat pan/square as the semantically clean loops.

6.3 Attention horizon vs. world persistence

If revisit failure were merely a *window-size* artifact, enlarging the attention horizon should repair it. Sweeping the KV read window $w \in \{3, \ldots, 18\}$ (6× context) at 270p with the baseline condition, mean revisit PSNR *drops* from 11.8 (w=3) to $\approx$ 9.0 and never recovers: more context admits more stale frames and helps in no case (raw values: appendix); w=4 at 480p leaves baseline ERC unchanged (outback-40 -0.58 vs. -0.77; outback-80 -0.36 vs. -0.33; scene 00, seed 0) and Anamnesis unaffected ($+3.12$/$+5.26$); w=6 exceeds 24 GB at 480p. The failure mode is *recollection*, not context capacity: the model does not use even the context it already has to reinstate places, which is exactly the gap our mechanism targets.



Figure 1: Revisit qualitative comparison on outback-40 (left block) and square-40 (right block): departure, pose-exact revisit, and novel-view control frames under the four conditions. Under Anamnesis (renoise), the regenerated revisit frame reinstates the remembered scene (buildings, horizon line, palette); the random control with more anchor firings does not, and the kv override matches baseline. Frames are uncurated, first-run outputs.

6.4 Ablations

Re-noising level (dose–response). Varying the corruption level of the anchor before its denoising pass yields a clean monotone curve in ERC-PSNR at scene 00, seed 0 (outback-40 / pan-45 / square-40 / outback-80); recall that the 4-step schedule is descending, so $t_3{\approx}321$ is the *low*-noise end and $t_1{\approx}821$ the high-noise end: $k^*{=}t_3$ (light corruption, one refinement step) gives $+3.25/+1.76/+1.55/+1.14$; $k^*{=}t_2$ decays roughly proportionally ($+2.61/+0.95/+0.58/+0.14$); $k^*{=}t_1$ (heavy corruption, three steps) is null ($-0.37/-0.83/+0.34/-0.03$; full sweep: appendix). The frozen prior is an active amnesiac: given enough sampling freedom it regenerates the past away. Recollection works by *constraining* the model's imagination: the memory must enter almost intact, with a single low-noise step left to adapt it to the revisited viewpoint. This mirrors the SDEdit sweet spot and is the operational form of "memory as constrained re-rendering."

Detection threshold. Tightening the translation gate from $0.15\,\text{m}$ to $0.05\,\text{m}$ leaves pan-45 and square-40 *identical* (pose-exact detection) and slightly improves outback-80 (12.46 vs. 12.39 dB, one precisely timed anchor), but zeroes outback-40: chunk-start pose sampling misses the corridor. Detection precision, not recall breadth, is the binding constraint, motivating swept-volume pose indexing as future work.

Is recollection just copying? A fair objection: retrieval-and- inject might merely paste a stored frame back, inflating similarity without memory. We test this offline (scene 00, seed 0) by comparing, per recollected rollout, the SSIM between its revisit frame and its own departure frame (a copy forces this toward 1). Ratios are 0.47 / 0.23 / 0.42 / 0.23 (outback-40 / -80 / pan-45 / square-40), statistically indistinguishable from the baseline's self-similarity (0.28 / 0.27 / 0.34 / 0.27): Anamnesis copies nothing. Against the *paste ceiling* ($1{-}\text{SSIM}(d,n)$, 0.64–0.73), short excursions close 17% / 9.5% of it and long excursions ≈ 0 on the SSIM scale while still gaining PSNR and LPIPS: the anchor re-biases geometry and palette, and texture is re-synthesized. The injection boundary is smooth at $\leq$93 frames ($\leq 1.04\times$ median frame-diff) and costs visible discontinuity at 177 (2.4–$3.1\times$), the honest price of a single-slot, in-band re-render, which multi-chunk anchoring should amortize.

Teleport-restart is not recollection either. The natural retrieval baseline re-initializes generation from memory: condition on the anchor frame as a fresh first frame and replay the final $k{=}8$ actions. It reproduces the given frame faithfully (mean abs. error 0.04–0.05), yet its revisit ERC is *negative* on all four loops (-1.64 to -0.20) because the fresh segment moves 1.4–$2.3\times$ the baseline median frame-diff, re-animating the anchor *away* from the remembered place. Recollection must enter *in-stream*, where the ongoing context constrains the re-rendering, precisely the channel Anamnesis uses and the KV null result (finding 3) isolates.

Recollection curves. Fig. 2 (top) traces ERC across anchor corruption levels; (middle) sweeps gate-pose jitter ($\sigma_t \in \{0, 0.1, 0.3\}$ m); (bottom) plots ERC against excursion length on the four-

point 270p grid. The dose–response replicates on scene 01: ERC rises monotonically with anchor cleanliness on three of the four loops (outback-40 $0.98 \to 3.91 \to 5.01$; pan-45 $-2.71 \to -0.26 \to 0.99$; square-40 $0.98 \to 1.33 \to 1.49$ dB from t_1 to t_3); on outback-80 the cleanest anchor *dips* ($0.76 \to 1.72 \to 1.55$), the cost of stretching a single anchor across a long excursion (see caveat 2). The gate degrades gracefully rather than collapsing: mean ERC falls $2.27 \to 1.20$ dB at σ_t=0.1 m and $\to 0.49$ dB at 0.3 m, with two loops still positive under mild noise. Recollection stays above the baseline at every excursion length tested (up to 323 frames), though the margin is non-monotonic at this single-seed 270p probe and we claim no decay law.

7 Discussion

The experiments give the phenomenological frame empirical teeth. **Retention and protention are architecturally present; recollection is not.** The KV cache retains the immediate past and the model anticipates its continuation, yet under the baseline the frozen model's excess over the novel-view control is negligible ($+0.35$ dB; *negative* on the pan loop, and positive only where prior typicality or corridor geometry supplies self-similarity for free): asked to reinstate a place, it returns what its prior would have invented anyway. This is the Husserlian gap made measurable.

Memory as constrained re-rendering. The dose–response of Sec. 6.4 carries the argument. Recollection fails not because the content is absent (the anchor latent is byte-identical to what was previously generated) but because the frozen prior, left with sampling freedom, regenerates the past into its own typicality: a heavily corrupted anchor (k^*=t_1, three steps) is erased, while a barely corrupted one (k^*=t_3, one step) survives and is re-rendered at the revisited viewpoint. Remembering, in a generative world model, is thus an act of *restraint*: the system must be prevented from re-imagining the past into the present. This is the inverse of the retrieval intuition. It explains why the KV-override control, which injects the memory *perfectly intact* into the retention channel, does nothing: intactness in the wrong channel is forgotten wholesale; only near-intactness in the generative channel (one denoising step from noise to perception) constitutes recollection. Both halves of this finding have independent witnesses: VRAG (Chen et al., 2025) observed from the training side that video models ignore naively concatenated clean retrieved frames and must be trained with reduced conditioning noise to use them. We show the same variable is the whole story at inference, with no training; and Li (2026) observed in LLMs that surgically overridden KV entries are ignored, the same channel asymmetry measured here in a video DiT.

Identity without substance. Loop closure operationalizes the Ship of Theseus: the world model rebuilds the scene from a barely-corrupted trace, and we ask only whether the rebuilt world is the *same place* by pose-exact comparison. That place identity survives regeneration at all ($+2.24$ dB mean ERC-PSNR) while instance appearance partially does not (semantic excess mixed, pooled $p = 0.069$) is itself the thesis: the identity of a place is carried by its configuration, not its matter.

8 Limitations

(i) All recollection episodes use *self-generated* anchors; the buffer can compound the model's own distortions. (ii) Detection consumes exact camera poses; free-form exploration needs SLAM-quality tracking, and chunk-start sampling is coarse (0.6 m). Swept-volume indexing would fix the gate-0.05 recall loss. (iii) The evidence pool (4 scenes $\times$ 2 seeds) is small, drawn from the release's example frames; broader scene sampling and cross-model replication are open. (iv) k^* and the gates are hand-set; a closed-loop controller over recollection strength is untested. (v) DINOv2 responds fastest to layout and palette; instance-level and 3D-consistency evaluation (point-cloud reprojection) is left open.

9 Conclusion

We identified the recollection gap in autoregressive world models, gave it a phenomenological formulation, and repaired it at inference time: an episodic buffer, a pose-gated detector, and one low-noise refinement step recover a measurable part of lost place memory on a frozen world model, at zero training cost. Multi-chunk and swept-volume indexing should extend recollection to longer excursions and free exploration. We release our benchmark, metrics, and wrapper code.

REPRODUCIBILITY STATEMENT

All experiments run on a single node with 4×RTX-4090 (24 GB) from the publicly released LingBot-World-Fast checkpoint; no training is involved. The supplement contains the full wrapper and benchmark code (episodic buffer, pose-gated detector, both injection mechanisms), the trajectory DSL compiler and all loop definitions, per-run metrics for the complete run matrix (four scenes × two seeds × four conditions plus ablation sweeps), the offline metric implementation (PSNR/SSIM/LPIPS/DINOv2), and all generated videos. Every hyperparameter (gates, margins, anchor-noise index, seeds, window sizes) is listed in Appendix A and its launch scripts in the released code.

ETHICS STATEMENT

Persistent scene memory in world models could support surveillance-style reconstruction of visited environments. The mechanism studied here only re-renders what the model itself already generated during its own rollout, adds no new perception, and was evaluated exclusively on synthetic example scenes; we see no additional risk relative to the underlying world models.

REFERENCES

Mahault Albarracin, Riddhi J. Pitliya, Maxwell J. D. Ramstead, and Jeffrey Yoshimi. Mapping husserlian phenomenology onto active inference. *arXiv preprint arXiv:2208.09058*, 2022.

Anonymous. Reworld: An interactive world model with long-horizon memory. *arXiv preprint arXiv:2608.23565*, 2026.

Adrià Puigdomènech Badia, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Timple, David Tracey, Xavier Glorot, Georg Ostrovski, Rémi Munos, et al. Never give up: Learning directed exploration strategies. *ICLR*, 2020.

Charles Blundell, Benigno Uria, Alexander Pritzel, Yazhe Li, Avraham Ruderman, Joel Z Leibo, Jack Rae, Daan Wierstra, and Demis Hassabis. Model-free episodic control. *arXiv preprint arXiv:1606.04460*, 2016.

Jake Bruce, Michael Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Bai, Adil Rockenbauer, Yi Bai, et al. Genie: Generative interactive environments. *ICML*, 2024.

Taiye Chen, Xun Hu, Zihan Ding, and Chi Jin. VRAG: Learning world models for interactive video generation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Decart and Etched. Oasis: A universe in a transformer. `https://oasis.decart.ai`, 2024.

DeepMind Team. Genie 3: A new frontier for world models. *DeepMind Research Blog*, 2025.

Y. Deng et al. Worldscore: A unified evaluation benchmark for world generation. *arXiv preprint arXiv:2504.00983*, 2025.

Y. Du et al. R2m-bench: Revisit memory benchmark for world models. *arXiv preprint arXiv:2608.27328*, 2026.

Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O Stanley, and Jeff Clune. First return, then explore. *Nature*, 590:580–586, 2021.

Z. Fan et al. Wbench: A comprehensive multi-turn benchmark for interactive video world model evaluation. *arXiv preprint arXiv:2605.25874*, 2026.

Ruili Feng, Han Zhang, Zhantao Yang, Jie Xiao, Zhilei Shu, Zhiheng Liu, Andy Zheng, Yukun Huang, Yu Liu, and Hongyang Zhang. The matrix: Infinite-horizon world generation with real-time moving control. *arXiv preprint arXiv:2412.03568*, 2024.

Klaus Greff, Sjoerd van Steenkiste, and Jürgen Schmidhuber. On the binding problem in artificial neural networks. *arXiv preprint arXiv:2012.05208*, 2020.

Yixuan He et al. Matrix-game: Interactive world generation. *arXiv preprint arXiv:2506.18701*, 2025a.

Yixuan He et al. Matrix-game 2.0: Real-time and controllable interactive world model. *arXiv preprint arXiv:2508.13009*, 2025b.

Yixuan He et al. Matrix-game 3.0: Real-time and streaming interactive world model with long-horizon memory. *arXiv preprint arXiv:2604.08995*, 2026a.

Yixuan He et al. Matrix-game 3.5: Enhancing real-time streaming interactive world models with patch memory. *arXiv preprint arXiv:2608.29910*, 2026b.

Z. Hu et al. Alaya-evoke: From linear-scaling supervision to endless world. *arXiv preprint arXiv:2608.13546*, 2026.

Junchao Huang, Xinting Hu, Boyao Han, Shaoshuai Shi, Zhuotao Tian, Tianyu He, and Li Jiang. Memory forcing: Spatio-temporal memory for consistent scene generation on minecraft. *arXiv preprint arXiv:2510.03198*, 2025.

Edmund Husserl. *The Phenomenology of Internal Time-Consciousness*. Indiana University Press, 1964.

Bojie Li. Models take notes at prefill: KV cache can be editable and composable. *arXiv preprint arXiv:2606.17107*, 2026.

Yulin Li et al. Hunyuan-gamecraft: High-dynamic interactive game video generation with hybrid hierarchical timestep sampling. *arXiv preprint arXiv:2506.17201*, 2025.

Xiaofeng Mao et al. Yume: An interactive world generation model. *arXiv preprint arXiv:2507.17744*, 2025.

Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jianjun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. In *ICLR*, 2022.

Sodtavilan Odonchimed, Tatsuya Matsushima, Simon Holk, Yusuke Iwasawa, and Yutaka Matsuo. Retrieve-augmented generation for speeding up diffusion policy without additional training. *arXiv preprint arXiv:2507.21452*, 2025.

S. Park et al. Echo-memory: A controlled study of memory in world models. *arXiv preprint arXiv:2606.09803*, 2026.

Alexander Pritzel, Benigno Uria, Sriram Srinivasan, Adrià Puigdomènech Badia, Oriol Vinyals, Demis Hassabis, and Charles Blundell. Neural episodic control. *ICML*, 2017.

Maxwell JD Ramstead et al. Mapping husserlian phenomenology onto active inference. *arXiv preprint arXiv:2208.09058*, 2022.

NVIDIA Research et al. Sana-wm: Efficient minute-scale world modeling with hybrid linear diffusion transformer. *arXiv preprint arXiv:2605.15178*, 2026.

Robbyant Team. Advancing open-source world models. *arXiv preprint arXiv:2601.20540*, 2026.

Y. Shi et al. Worldmark: A benchmark for world memory. *arXiv preprint arXiv:2604.21686*, 2026.

Vighnesh Subramaniam, Colin Conwell, Boris Katz, Andrei Barbu, and Brian Cheung. Network of theseus (like the ship). *arXiv preprint arXiv:2512.04198*, 2025.

Zhihao Tu et al. Worldplay: Towards long-term geometric consistency for real-time interactive world modeling. *arXiv preprint arXiv:2512.14614*, 2025.

Dani Valevski, Yaniv Leviathan, Moab Arar, and Shlomi Fruchter. Diffusion models are real-time game engines. *arXiv preprint arXiv:2408.14837*, 2024.

Rishi Veerapaneni, John D. Co-Reyes, Michael Chang, Michael Janner, Chelsea Finn, Jiajun Wu, Joshua Tenenbaum, and Sergey Levine. Entity abstraction in visual model-based reinforcement learning. *arXiv preprint arXiv:1910.12827*, 2019.

Z. Wang et al. Dreamx-world 1.0: A general-purpose interactive world model. *arXiv preprint arXiv:2606.16993*, 2026a.

Z. Wang et al. Mind: Benchmarking memory consistency and action control in world models. *arXiv preprint arXiv:2602.08025*, 2026b.

Xindi Wu, Sven Elflein, James Lucas, Olga Russakovsky, Laura Leal-Taixé, Despoina Paschalidou, Jonathan Lorraine, and Aljoša Ošep. Addressable memory for video world models. *arXiv preprint arXiv:2608.07408*, 2026.

Guangxuan Xiao, Yuandong Tian, Beidi Chen, Song Han, and Mike Lewis. Efficient streaming language models with attention sinks. In *ICLR*, 2024.

Xiao Xu et al. Worldmem: Accurate and consistent world simulation with memory. *arXiv preprint arXiv:2504.12369*, 2025.

Qinyu Yang, Haoxin Chen, Yong Zhang, Menghan Xia, Xiaodong Cun, Zhixun Su, and Ying Shan. Noise calibration: Plug-and-play content-preserving video enhancement using pre-trained video diffusion models. In *European Conference on Computer Vision (ECCV)*, 2024.

Tianyi Zeng, Jincheng Gao, Tianyi Wang, Zijie Meng, Miao Zhang, Jun Yin, Haoyuan Sun, Junfeng Jiao, Christian Claudel, Junbo Tan, and Xueqian Wang. When to lock attention: Training-free KV control in video diffusion. *arXiv preprint arXiv:2603.09657*, 2026.

Wenqu Zhao, Chi Xuemin, Xin Zhang, Guoqing Ma, Baorun Li, Jianjie Fang, Peizhi Tang, Chen Gao, and Wei Wu. DensityKV: Density-guided KV cache compression for long video generation. *arXiv preprint arXiv:2608.27922*, 2026.

A Implementation Details

Wrapper. `WanI2VFastAnamnesis` subclasses the released `WanI2VFast` and only interposes at chunk boundaries: after chunk c is denoised to x_0, the buffer stores (c, x_0, P_{3c}). Under mode *renoise*, before generating a chunk the detector queries the buffer with that chunk's own pose P_{3c}; on a hit the chunk's initial noise is replaced by $\mathrm{add_noise}(x_0^{\mathrm{anc}}, \epsilon; t_{k^*})$ and only timesteps $k^*..3$ are executed (fewer steps, since recollection is *cheaper* than synthesis). Under mode *kv*, at the end of chunk c the detector queries with the *next* chunk's pose $P_{3(c+1)}$; on a hit, chunk c's context-write (the $t{\approx}0$ forward that pushes x_0 into the KV cache) writes the *anchor* latent instead of $x_0^{(c)}$, so chunk $c{+}1$ attends to recollected content as its recency slot. Mode *random* fires the same re-noising machinery on a uniformly random eligible anchor; mode *none* is the untouched pipeline.

KV cache. Per-layer caches $\{k, v, \mathrm{global_end}, \mathrm{local_end}\}$ hold $[1, w \cdot s_f, 40, 128]$ bf16 entries, where s_f is the per-frame token count (1508 at 480p). The read window is the last $\min(\mathrm{max_attn}, \mathrm{local_end})$ tokens with left-shift eviction preserving sink tokens; the horizon sweep varies this read window.

Trajectory DSL. Action strings (`w`-forward, `a`/`d`-strafe, `s`-back, `i`/`k`-pitch, `j`/`l`-yaw, `none`-hold; unit: frames at 0.05 m and 2° per frame) compile to exact SE(3) pose lists; loop closures are analytic ($2{\times}10^{-16}$ m translation error), so revisit frame pairs are ground-truth aligned without any perception.

Memory engineering on 4×24 GB. The released checkpoint is fp32 (69 GiB); under FSDP's `MixedPrecision` this resides fp32 and OOMs, so we cast to bf16 before sharding (9.1 GiB/rank steady). At 480p the KV window $w{=}3$ costs 3.7 GiB; $w{=}18$ at 270p (8.4 GiB) is the sweep ceiling. All experiments ran on a single node without model offloading.

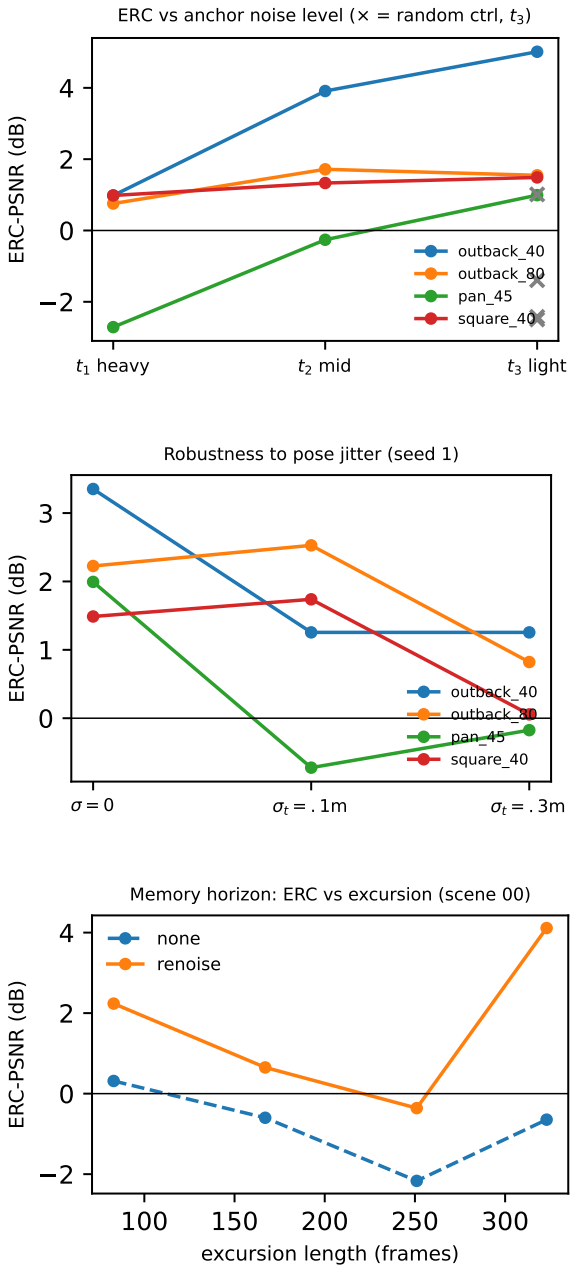

Figure 2: Recollection curves (ERC-PSNR). **Top:** anchor re-noising level (scene 01): light corruption with one refinement step preserves; heavy corruption is regenerated away. **Middle:** gate-pose jitter (scene 00, seed 1): degradation is graceful: mean ERC halves at σ_t=0.1 m and approaches zero at 0.3 m. **Bottom:** excursion length (scene 00, 2 seeds, 270p).

B Additional Results

Horizon sweep raw values (mean revisit PSNR over loops, 270p, none): w=3: 11.77; w=6: 9.04; w=9: 8.93; w=12: 8.98; w=18: 9.11. Per-loop values and ERC columns are released with the benchmark code.

DINO similarity table. Table 2 reports the DINOv2 (ViT-B/14) cosine similarities and their excess, for the same grid as Table 1.

Loop	Cond	DINO	ERC-DINO	n
outback_40	none	0.353 ± 0.094	-0.106 ± 0.099	(8)
outback_40	random	0.370 ± 0.195	-0.327 ± 0.321	(8)
outback_40	renoise	0.674 ± 0.155	0.219 ± 0.186	(8)
outback_40	kv	0.356 ± 0.097	-0.102 ± 0.093	(8)
outback_80	none	0.408 ± 0.072	0.020 ± 0.049	(8)
outback_80	random	0.466 ± 0.242	0.120 ± 0.298	(8)
outback_80	renoise	0.315 ± 0.063	-0.080 ± 0.086	(8)
outback_80	kv	0.397 ± 0.074	0.011 ± 0.053	(8)
pan_45	none	0.375 ± 0.073	0.002 ± 0.064	(8)
pan_45	random	0.495 ± 0.257	-0.016 ± 0.069	(8)
pan_45	renoise	0.699 ± 0.070	0.321 ± 0.041	(8)
pan_45	kv	0.367 ± 0.094	-0.010 ± 0.081	(8)
square_40	none	0.514 ± 0.066	0.120 ± 0.054	(8)
square_40	random	0.463 ± 0.243	0.153 ± 0.284	(8)
square_40	renoise	0.278 ± 0.060	-0.121 ± 0.078	(8)
square_40	kv	0.518 ± 0.073	0.125 ± 0.053	(8)

Table 2: Semantic (DINOv2) revisit similarities and excess for the main grid. ERC-DINO is the excess over the novel-view control.

Failure cases. Recollection fails when (i) the excursion visits genuinely *new* geometry (novel-view control catches this; ERC correctly ≈ 0), (ii) the scene changed dynamically between visits (anchor regeneration reproduces the stale scene), and (iii) anchors are sparse (single-anchor loops recover less than corridor-revisit loops, though still $+1$ dB).