NEURIPS 2026

AdaState

Self-Evolving Anchors for Streaming Video Generation

Yusuf Dalva · Pinar Yanardag

Virginia Tech · adastate.github.io

AdaState generations



THE PROBLEM

Streaming video stays stuck on its first frame

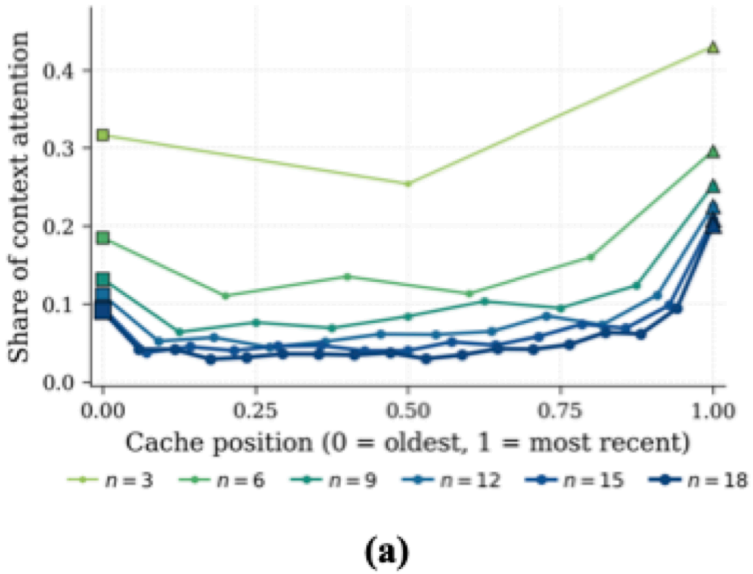


Far beyond the training horizon. *"A dynamic FPV aerial flight through a vibrant underwater coral city…"*

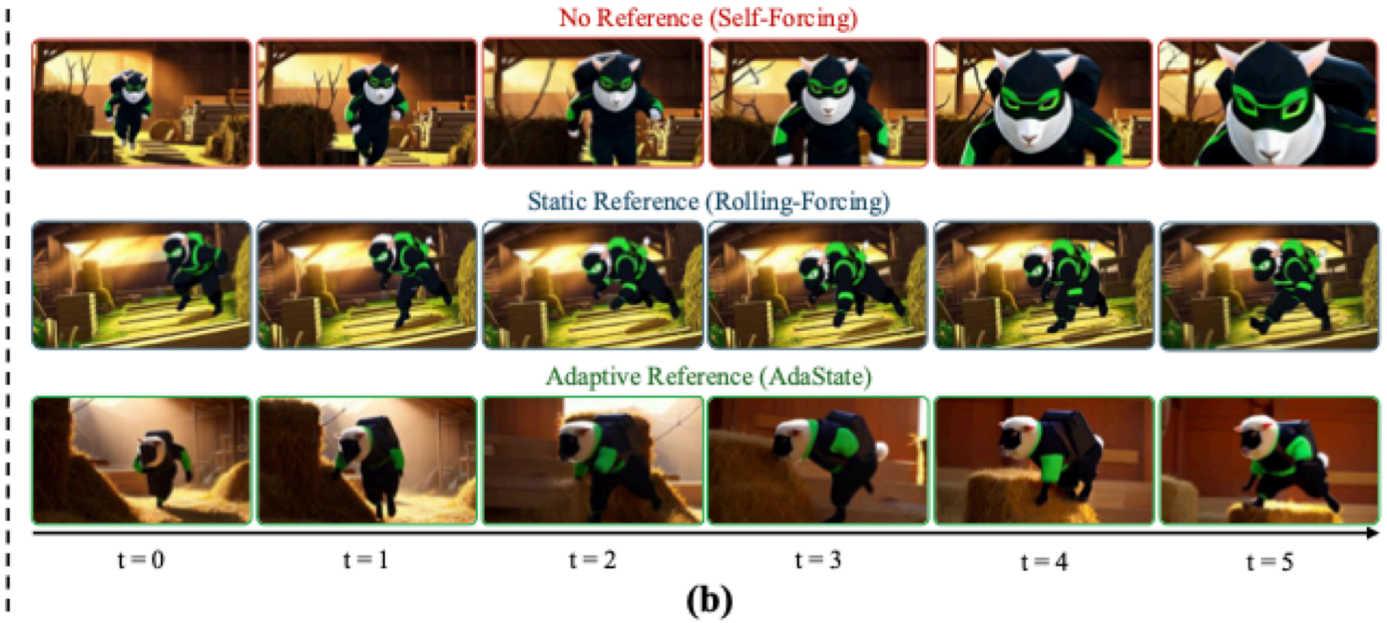
- **Anchored by design.** The first frame's KV sits in the first cache slot as the scene reference.
- **Too attractive.** It is the cleanest entry in the cache, so attention piles onto it.
- **Frozen scenes.** Motion and scene progression are traded for static consistency.

ANALYSIS

Two cache slots dominate, and one is frozen



(a)



(b)

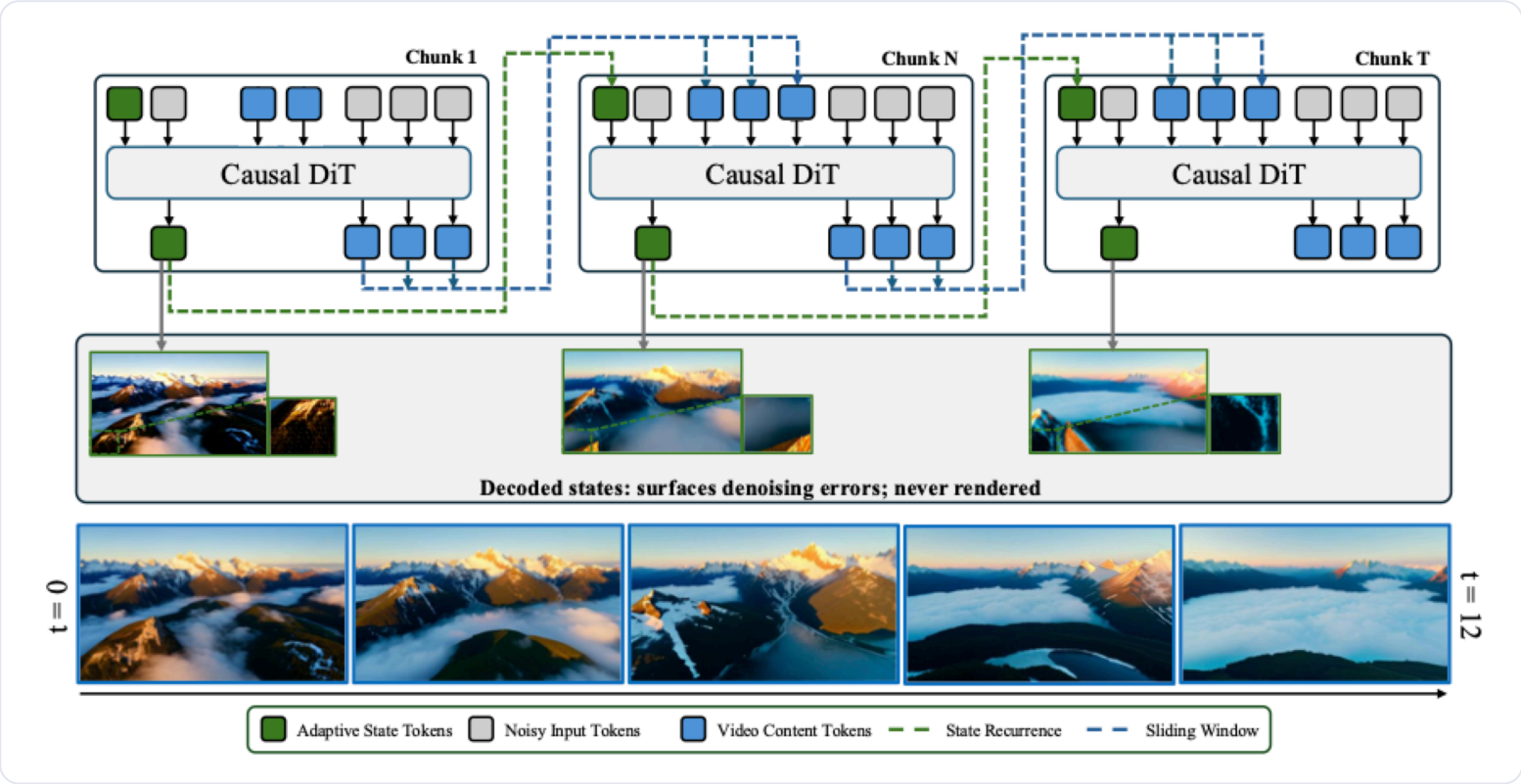
Bimodal attention. The first slot and the newest frame take most of the attention.

Flat middle. Everything in between gets roughly uniform attention.

Our lever. Keep the attention pattern; change what sits in the slot the model trusts most.

METHOD

AdaState: the model writes its own anchor



- **Hidden state.** One extra latent frame is denoised with each chunk, starting from noise, and never rendered.
- **Write to the anchor slot.** Its clean KV replaces the anchor; content goes to the sliding window.
- **Built-in recurrence.** The denoiser is the transition function and the KV cache is the carrier. No new modules.

TRAINING

Horizon weighting: train where the state matters

$$\mathcal{L} = \frac{1}{2} \sum_{i=0}^{N-1} w_i \left\| \hat{x}_i - \mathrm{sg}\left[\hat{x}_i - \frac{S_{\text{fake}} - S_{\text{real}}}{\gamma} \right] \right\|^2$$

per-frame weight
the only change to the loss

standard DMD target
same as Self-Forcing

$$w_i = 1 + \alpha \cdot \frac{i}{N-1}$$

slope α
how strongly late frames count

frame position
grows from the first frame to the last

Why late frames?
They carry the most drift, and by then the first frame has left the window, so the state is their main reference.

No state loss.
The state learns only through the attention content pays to it.

The motion gain comes from the state. Weighting alone does not produce it.

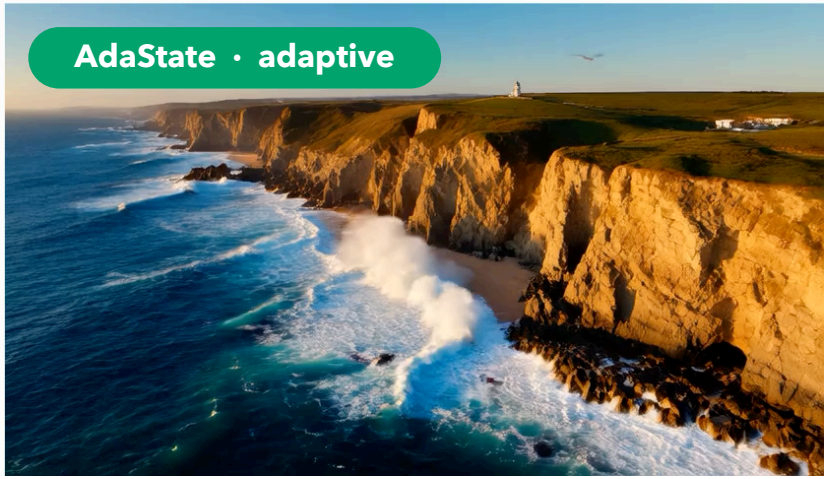
RESULTS

On long rollouts, AdaState keeps the scene moving









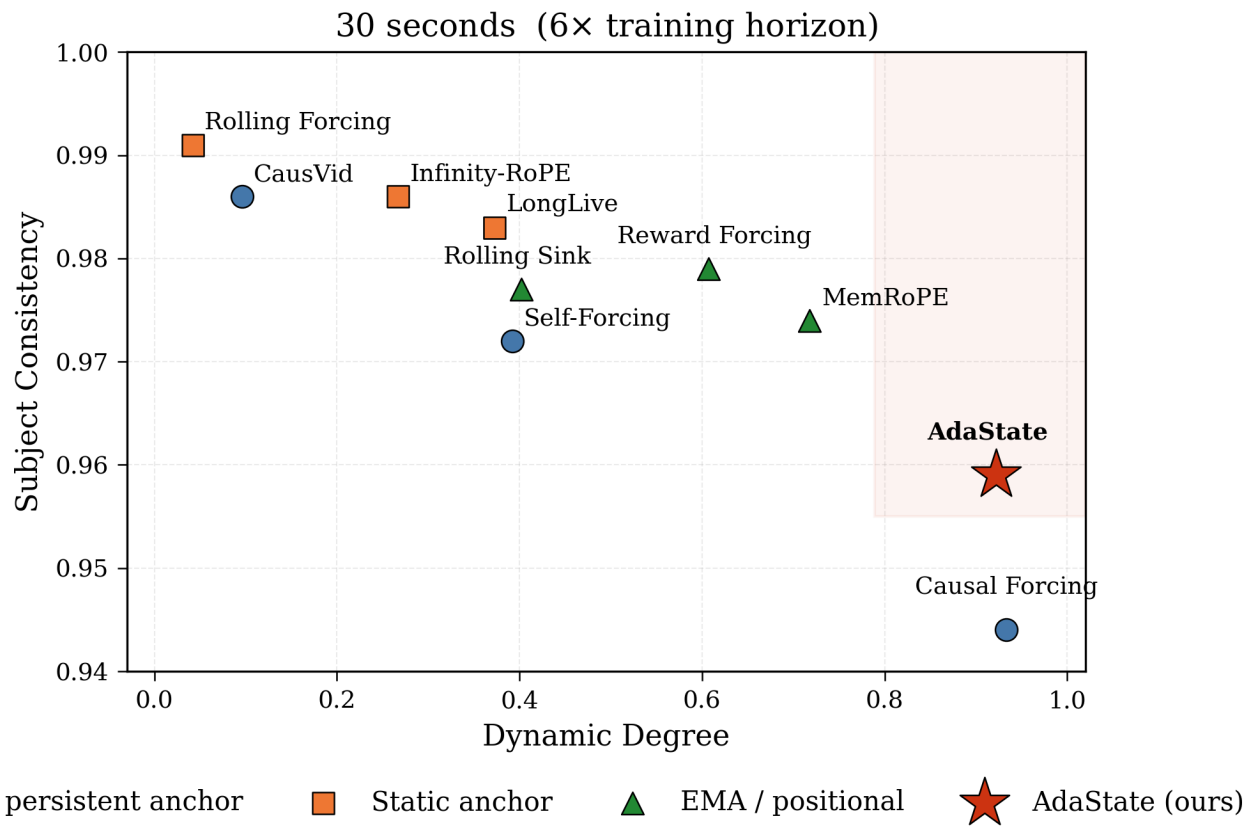
Far beyond the training horizon

"A sweeping aerial drone shot above dramatic coastal cliffs at golden hour…"

- **No anchor** drifts off the prompt into artifacts.
- **Static / EMA** anchors hold the same cliff throughout.
- **AdaState** travels the coast; new terrain keeps appearing.

RESULTS

AdaState breaks the consistency–dynamics tradeoff



- **Baselines trade one for the other.** Stable anchors freeze the scene; anchor-free models drift.
- **AdaState gets both.** It alone reaches high motion and high consistency together.
- **Best overall quality.** Top VBench and VisionReward scores, within and beyond the training horizon.

User study: people rate AdaState highest on coherent progression and prompt following.

SUMMARY

Takeaways

1

The anchor is the bottleneck

A frozen first frame at the most-attended cache slot suppresses dynamics.

2

Let the model write its own anchor

A hidden adaptive state, denoised with each chunk, evolves with the scene and keeps identity.

3

Denoising is a recurrence

No external memory: the generator updates the state and the KV cache carries it.

Videos and paper adastate.github.io