Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Doctoral Consortium Track

Context-Adaptive World Models for Visual Out-of-Distribution Generalization

Shubham Subhnil
School of Computer Science and Statistics, Trinity College Dublin
subhnils@tcd.ie

Abstract

A vision-based robot can be faced with changing friction, gravity, or motor behavior, whose effects are observable only from pixels mid-episode. Visual model-based reinforcement learning struggles in such settings because most world models assume the latent factors driving these changes stay fixed during interaction, or fold them into a single context embedding. My thesis is developing a context-adaptive Mamba-based world model that infers K context vectors from short pixel-trajectory windows. Each vector steers a disjoint slice of the Mamba backbone through block-diagonal feature-wise linear modulation (FiLM), so distinct dynamics signals travel along parallel pathways rather than collapse into a single channel. The policy trains on imagined rollouts conditioned on the inferred context, and I show that its suboptimality is bounded by $\sqrt{K}$ times the context-inference error along a Lipschitz chain through FiLM, the SSM, and the decoder. Evaluation spans within-episode shifts on DeepMind Control (Walker, Quadruped) and mode/difficulty switching on Atari, against an inferred-context baseline and a privileged-context oracle.

1 Introduction and Motivation

Out-of-distribution (OOD) generalization in pixel-driven, partially observable reinforcement learning is a hard online problem. Latent features of the environment such as friction, actuator strength or reward-allocation rules shape what the agent sees and earns but cannot be observed from any single frame; their trajectory-level effect is the agent's *context* [Benjamins *et al.*, 2023]. A robust agent must infer several such factors from pixels and stay calibrated when they shift at *different* times within an episode.

Existing world models leave this setting open. The Dreamer family [Hafner *et al.*, 2025] and recent state-space model (SSM)-based world models [Wang *et al.*, 2025; Dao and Gu, 2024] are context-agnostic. Pixel-space context-conditioned world models [Yang *et al.*, 2025; Röder *et al.*, 2025; Prasanna *et al.*, 2024] either assume fixed context per episode or rely on a privileged context channel. DOMINO [Mu *et al.*, 2022] infers multi-factor context but on featurized state. Meta-RL methods that admit within-episode drift [Chen *et al.*, 2022; Liang *et al.*, 2024] carry context as a single latent. No prior method jointly meets the three demands: (i) multi-factor context inference from pixels, (ii) within-episode asynchronous tracking, and (iii) propagation of the current context into the world model used for imagination-based policy training.

This work fills this gap. Architecturally, it splits the conditioning across K inferred vectors through a block-diagonal FiLM head, so each vector drives an independent dynamics pathway. Theoretically, it provides the first bound linking the context-inference error in this setting to policy suboptimality through the FiLM Lipschitz chain.

2 Problem Formalization

The setting is a piecewise-stationary multi-factor contextual POMDP. The true context is $u_t = (u_t^1, \ldots, u_t^F)$ with each u_t^f piecewise constant and switching asynchronously; the transition, reward, and observation models are parameterized by u_t, and the agent observes only pixels and scalar rewards. The inferred context $c_t = (c_t^1, \ldots, c_t^K)$ is a set of K vectors that summarize the latent trajectory at distinct temporal scales rather than estimate individual factors. OOD evaluation probes factor values, factor combinations and switch cadences outside training.

3 Research Contributions

(1) Multi-slot context inference from pixels. An encoder infers K context vectors from a short causal pixel window via cross-attention; each slot has its own adaptive lookback and a fixed geometric temporal scale, so vectors attend to tokens over different horizons. Three losses shape the representation: a soft-label temporal contrastive loss anchors slots across same-phase steps, a persist-or-jump transition regularizer with a learned hazard smooths the inferred trajectory, and a Hilbert-Schmidt regularizer that discourages cross-slot dependence. The design extends the multi-factor context inference of [Mu *et al.*, 2022] and the segment-aware inference of [Chen *et al.*, 2022; Liang *et al.*, 2024] to the pixel setting, unsupervised.

(2) Block-diagonal FiLM on a Mamba backbone. Context vectors steer a Mamba selective-SSM world model [Dao

and Gu, 2024; Wang *et al.*, 2025] through block-diagonal FiLM [Perez *et al.*, 2018] at two recurrence-boundary sites: the per-head log time-step δ_t governing temporal selectivity, and the output gate $\mathbf{g}_t$ governing output routing, replacing the encoder-input concatenation of [Röder *et al.*, 2025; Prasanna *et al.*, 2024]. The block-diagonal FiLM head assigns each vector to a disjoint subset of SSM heads, so vector k modulates only that subset and gradients through it update only vector k. A ReZero-initialized parallel pathway captures the state-context interaction that the FiLM head cannot express. Tanh-bounded scales and near-zero output initialization keep the recurrent state bounded under any context.

(3) Context-aware imagination with formal guarantees. A policy trains on imagined rollouts from replay-buffer anchors with context held fixed and appended to the policy input as a phase signal; asynchronous switching enters training through FiLM-conditioned δ_t on real trajectories. Three formal results ground the design: (i) FiLM at the chosen sites preserves the SSM eigenvalue bound for any context; (ii) the cost of a factored variational posterior over the F switch indicators is bounded by $(F-1)\,h(\bar{\lambda})$ nats, small at sparse switch rates; (iii) policy suboptimality is bounded by $\sqrt{K}$ times the context-inference error plus the world-model error along a Lipschitz chain through FiLM, the SSM and the decoder, rate-invariant in the switch cadence.

4 Evaluation Plan

Continuous control on DeepMind Control (Walker, Quadruped) uses within-episode asynchronous Bernoulli switching on physics factors and reward direction, with $F \in \{2, 3\}$. Discrete control on the Arcade Learning Environment uses episode-aligned mode/difficulty switching on Alien and BankHeist. OOD splits hold out factor values, joint compositional corners and switch cadences spanning $0.2\times$ to $4\times$ the training rate. Results are compared to DRAMA [Wang *et al.*, 2025] (matched-backbone, context-agnostic), DALI [Röder *et al.*, 2025] (inferred context, Dreamer family), and cRSSM [Prasanna *et al.*, 2024] (privileged-context oracle). Ablations isolate the multi-scale role of slots, the chosen K, δ_t versus $\mathbf{g}_t$ modulation, surplus-slot pruning at test time, and a capacity-matched control.

5 Current Status and Impact

I am in the final stage of my doctoral studies. The multi-slot conditioning above is realized on Mamba-2 and Mamba-3 SSM backbones: the slot context vectors match a privileged-context oracle on rare-switch joint shifts, extend gains to unseen switch cadences, and transfer from continuous DMC physics to discrete episode-boundary switching on Atari. Remaining work covers automated slot-count selection, the backbone-expressivity ceiling at joint compositional corners, and few-shot adaptation of the inferred context at evaluation time. The thesis argues that pixel-driven RL agents can be made robust to unseen non-stationarity without privileged context.

Acknowledgments

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Frontiers for the Future grant number 21/FFP-A/8957. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[Benjamins *et al.*, 2023] Carolin Benjamins, Theresa Eimer, Frederik Schubert, André Biedenkapp, Bodo Rosenhahn, Frank Hutter, and Marius Lindauer. Contextualize me – the case for context in reinforcement learning. *Transactions on Machine Learning Research*, 2023.

[Chen *et al.*, 2022] Xiaoyu Chen, Xiangming Zhu, Yufeng Zheng, Pushi Zhang, Li Zhao, Wenxue Cheng, Peng Cheng, Yongqiang Xiong, Tao Qin, Jianyu Chen, and Tie-Yan Liu. An adaptive deep RL method for non-stationary environments with piecewise stable context. In *NeurIPS*, 2022.

[Dao and Gu, 2024] Tri Dao and Albert Gu. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In *ICML*, 2024.

[Hafner *et al.*, 2025] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *Nature*, 2025.

[Liang *et al.*, 2024] Anthony Liang, Guy Tennenholtz, Chih-Wei Hsu, Yinlam Chow, Erdem Biyik, and Craig Boutilier. DynaMITE-RL: A dynamic model for improved temporal meta-reinforcement learning. In *NeurIPS*, 2024.

[Mu *et al.*, 2022] Yao Mu, Yuzheng Zhuang, Fei Ni, Bin Wang, Jianyu Chen, Jianye Hao, and Ping Luo. DOMINO: Decomposed mutual information optimization for generalized context in meta-reinforcement learning. In *NeurIPS*, 2022.

[Perez *et al.*, 2018] Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *AAAI*, 2018.

[Prasanna *et al.*, 2024] Sai Prasanna, Anna Googasian, and Franziska Meier. Dreaming of many worlds: Learning contextual world models aids zero-shot generalization. *arXiv preprint arXiv:2403.10967*, 2024.

[Röder *et al.*, 2025] Jannik Röder, Stefan Roth, and Jan Peters. Dynamics-aligned latent imagination in contextual world models for zero-shot generalization. In *NeurIPS*, 2025.

[Wang *et al.*, 2025] Wenlong Wang, Ivana Dusparic, Yucheng Shi, Ke Zhang, and Vinny Cahill. DRAMA: Mamba-enabled model-based reinforcement learning is sample and parameter efficient. In *ICLR*, 2025.

[Yang *et al.*, 2025] Yupei Yang, Biwei Huang, Fan Feng, Xinyue Wang, Shikui Tu, and Lei Xu. Towards generalizable reinforcement learning via causality-guided self-adaptive representations. In *ICLR*, 2025.