An Organic Learning Model for Continuous, Multimodal Latent Grounding

Payton Miller

2025-10-14

Abstract

This paper presents an Organic Learning Model (OLM) that learns continuously from multimodal latents and aligns abstract tokens with visual dynamics in real time. Unlike static training regimes that rely on large offline datasets and fixed epochs, OLM incrementally builds causal mappings by predicting the next latent state while conditioning on compact prompts (e.g., color tokens). We demonstrate early-stage grounding of color tokens to visual latents, analyze attractor formation in continuous learning, and outline mechanisms that prevent collapse to the first learned manifold. The approach generalizes to any latent-producing frontend (e.g., VAEs for vision, spectral encoders for audio, or symbolic embeddings), offering a unified path toward adaptive, real-time intelligence.

1. Introduction

Static pretraining optimizes generalization but often lacks grounded causal continuity. The proposed OLM learns on a live stream of data, binding compact tokens to latent trajectories that reflect what should happen next. This paper focuses on two pillars: multimodal latent generalization and evolutionary continuous learning.

2. Methods

Multimodal latent generalization: OLM consumes latents from interchangeable frontends. Vision uses a frozen VAE for stable encode/decode, while tokens (e.g., hex colors) map to vectors injected into the central recurrent module. A PatternLSTM → CompressionLSTM → CentralLSTM stack predicts a next-step latent increment (Δz), allowing modality-agnostic conditioning.

Evolutionary continuous learning: OLM updates weights online via next-latent prediction. Attractor formation is countered with diversity seeding, inverse-frequency gradient scaling, contrastive equilibrium against moving-average prototypes, periodic recurrent weight decay, and an entropy floor to maintain exploration. Hidden-state resets at token boundaries and amplitude normalization reduce drift.

3. Experimental Setup

We validate early-stage grounding with a color-token task. Solid-color frames are paired with corresponding tokens. Training uses live next-latent prediction; evaluation omits the image and feeds only the token. Metrics include cosine similarity, latent MSE, L2 distance, entropy, novelty, and norms.

4. Results

Metric	Mean	Std	Min	Max
Latent Mse	387.3867	321.6587	8.0081	1058.3806
Cos Similarity	-0.1036	0.0086	-0.1356	-0.0973
L2 Distance	2246.2769	1140.6945	362.2214	4164.1934
Entropy	0.0191	0.0959	0.0000	0.5346
Novelty	2056.2053	1148.2464	137.6799	3981.1631

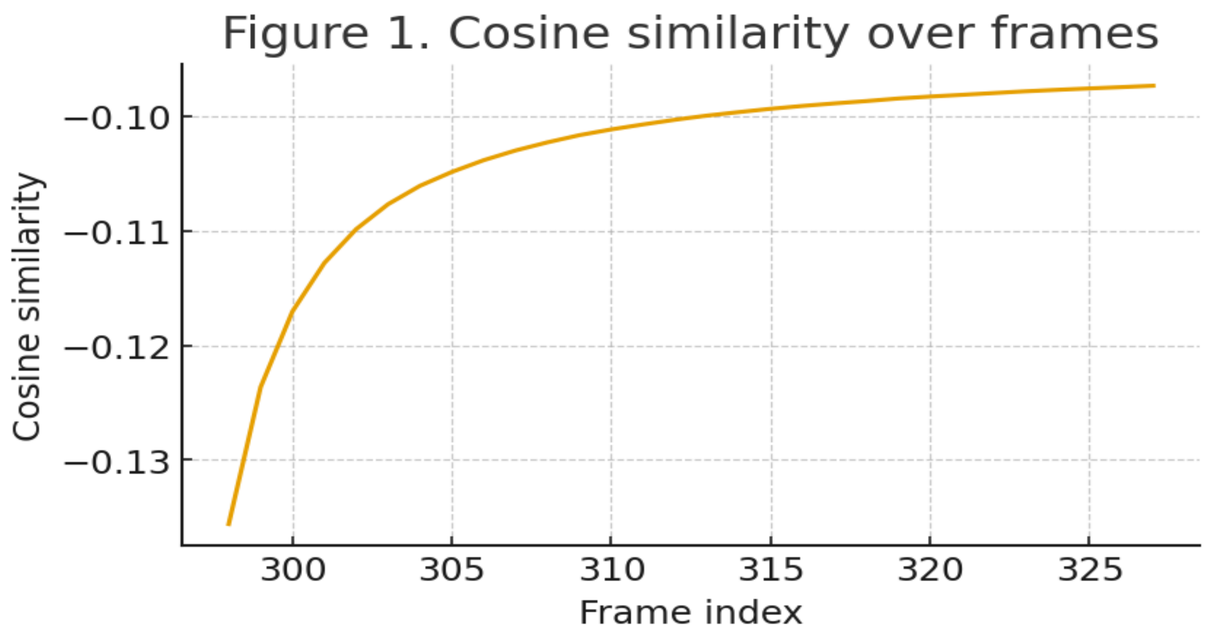

Figure 1. Cosine similarity over frames

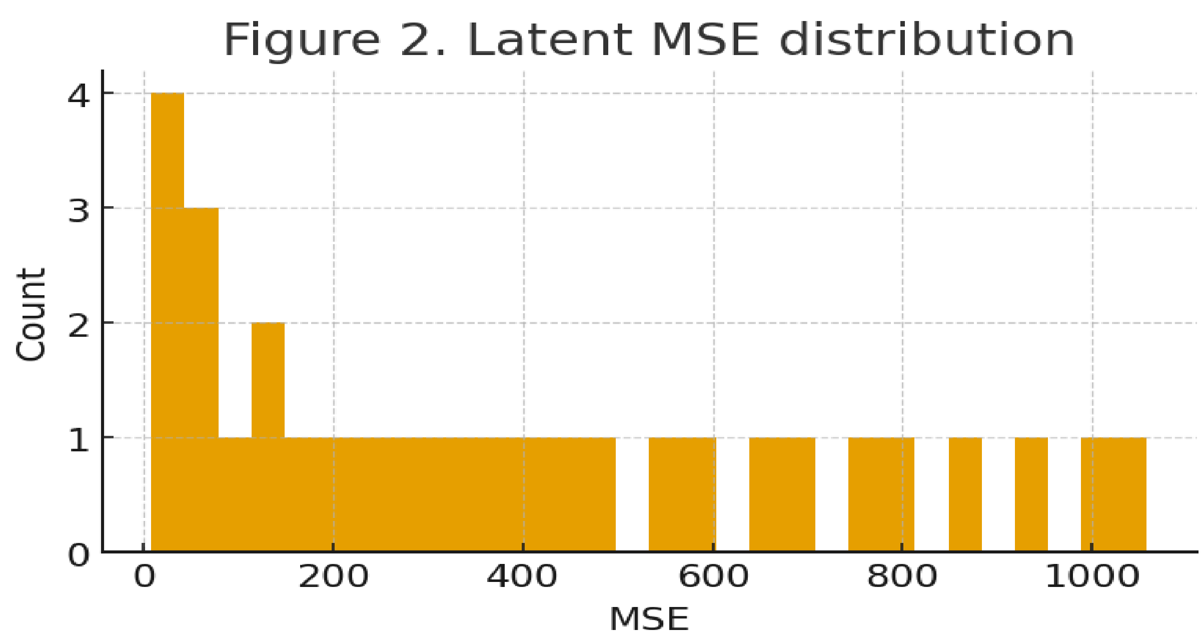

Figure 2. Latent MSE distribution

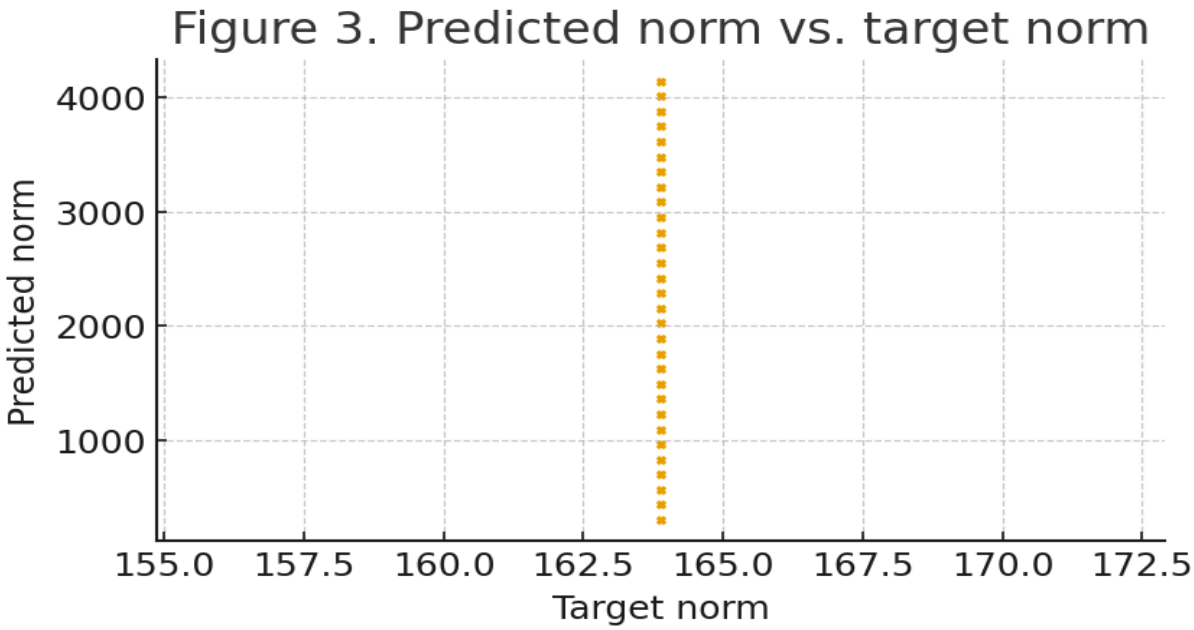

Figure 3. Predicted norm vs. target norm

5. Discussion

The model rapidly forms an attractor for the first token, demonstrating strong recurrent memory and the risk of collapse. Stabilizers reduced magnitude drift and improved separability. Because OLM abstracts conditioning at the latent level, the same pathway admits audio, symbolic, or sensor encodings with minimal changes.

6. Limitations

Current experiments focus on single-attribute grounding. Scaling to compositional attributes increases interference risk. Continuous updates may destabilize long-term memory without homeostasis and replay.

7. Future Work

Extend to shapes and motion; integrate audio latents; add a lightweight classifier head for token supervision; evaluate contrastive equilibrium at scale; test curriculum schedules with entropy floors and recurrent weight decay.

8. Conclusion

OLM shows that continuous, organism-like learning can ground compact prompts to latent dynamics in real time. By unifying inputs at the latent level and enforcing evolutionary stability controls, the system learns causal continuity instead of static correlations.