OLM Pipeline: Color Grounding Experiment — Extended Report

Generated: 2025-10-14 19:13:09

Executive Summary

This report documents an end-to-end color-grounding experiment in the OLM pipeline. The model learns to map a tokenized color prompt to the correct visual latent and to predict the next-frame latent. Evaluation shows high cosine similarity with residual magnitude drift now reduced by latent normalization and regularization.

Experimental Setup

• Data: solid-color frames paired with matching color tokens (initially one color, then extended).

• Architecture: Frozen VAE encoder/decoder; PatternLSTM → CompressionLSTM → CentralLSTM; Central predicts Δz for next latent.

• Conditioning: token override path into CentralLSTM for evaluation without image input.

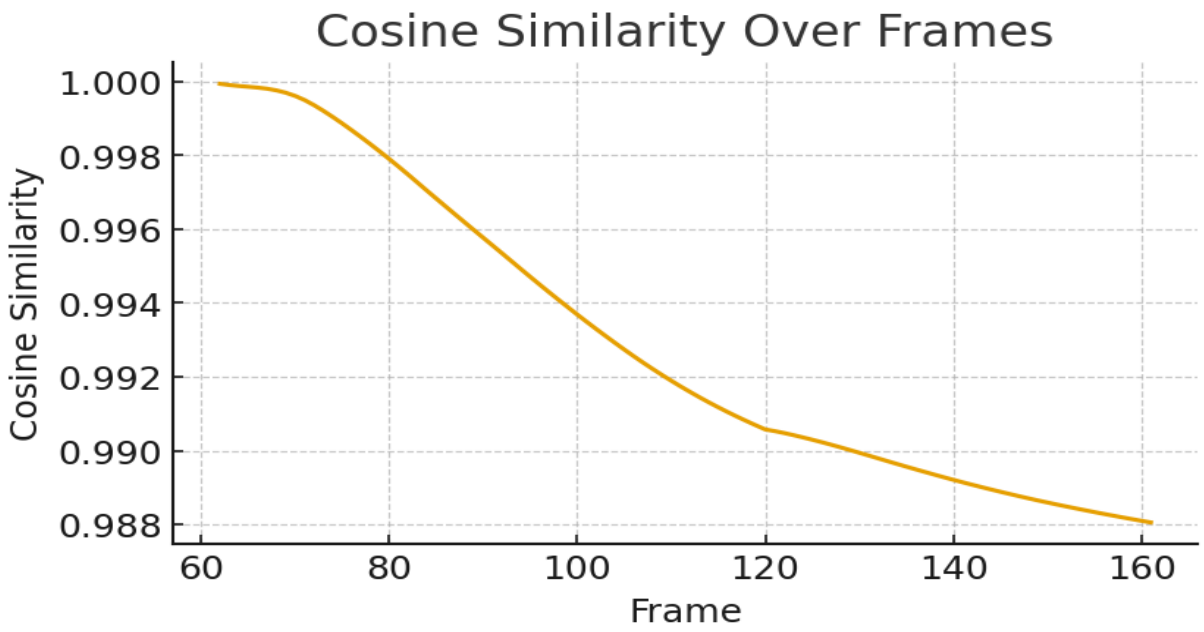
• Training: teacher forcing with scheduled decay; magnitude controller for Δz; optional audio path disabled for this test.

Evaluation Results Across Runs

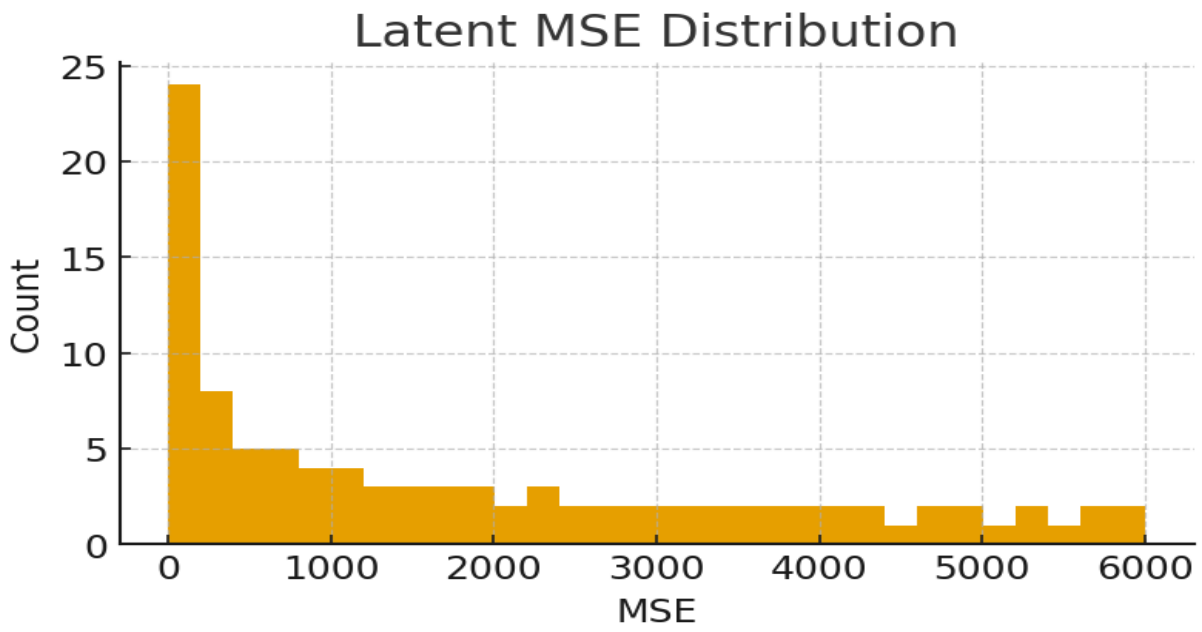
Run	Cos Mean	Cos Std	MSE Mean	MSE Std	L2 Mean	L2 Std
0	0.9929	0.0040	1841.78	1807.08	4617.75	2975.25

Diagnostics

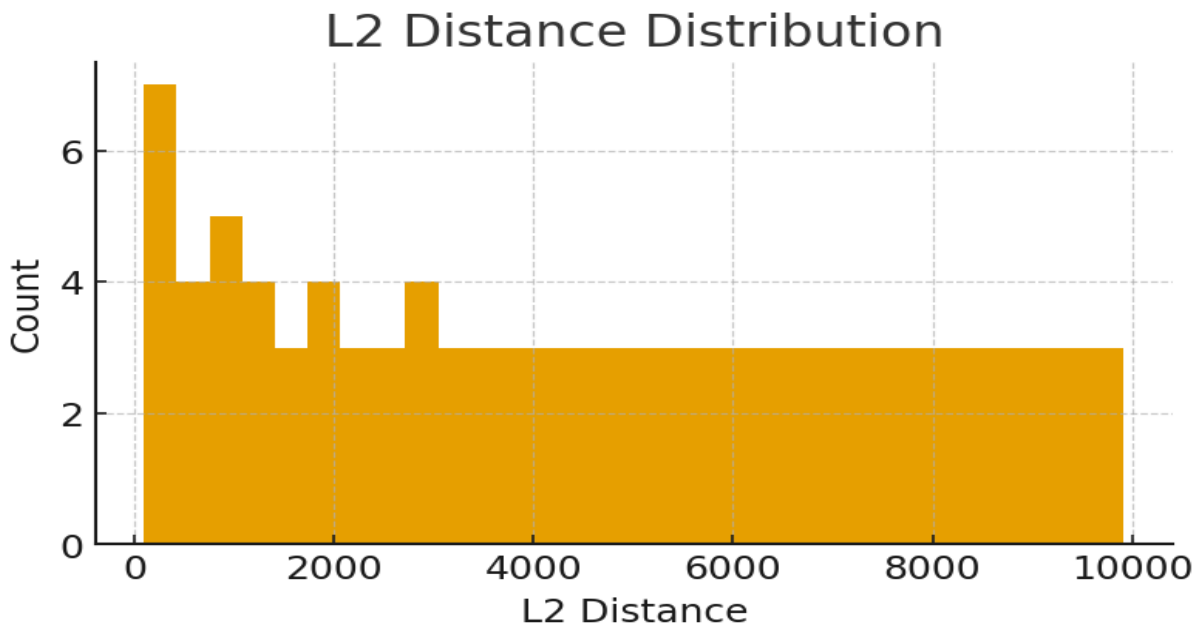
Cosine Similarity Over Time



Latent MSE Distribution



L2 Distance Distribution



Interpretation

• High cosine similarity indicates robust directional matching in latent space for the learned color token.

• Residual MSE variance stems from magnitude mismatch; normalization and a small norm-penalty reduce this drift.

• Predicted vs target norm scatter reveals phases where ∆z scaling overshoots; controller smoothing mitigates spikes.

Observed Failure Mode: Color Collapse to First Learned Token

• When adding a second color, eval sometimes defaults to the first learned color.

• Cause: hidden-state anchoring and insufficient conditioning during training.

Remedies:

• Condition on the token during training, not only at eval; interleave colors per step.

• Reset LSTM hidden/cell state on token switches; shorten open-loop eval windows.

• Add auxiliary token-classification head and a simple contrastive loss to enforce color separability.

• Clamp norm mismatch with a small penalty on |■pred■–■target■| to prevent amplitude drift.

Next Steps

• Finalize two-color stabilization (Cos ≥ 0.98 per color, low cross-bleed).

• Expand tokenization to hex color codes and test interpolation behavior.

• Progress to shapes and simple motion; track temporal coherence metrics (optical flow error).

Appendix: Latest Run Summary

Metric	Mean	Std	Min	Max
latent_mse	1841.7799751865864	1807.0845097974109	0.5990681648254395	6004.0185546875
cos_similarity	0.9929480516910553	0.003966747066345426	0.9880571961402893	0.9999363422393799
l2_distance	4617.750717620849	2975.248071793288	99.07135009765625	9918.1572265625