Mapping and Manipulating Latent Space through Structured Organic Learning Models (OLM): Toward Algebraic Control of Representations

Abstract Latent spaces in deep generative models are typically treated as black boxes: powerful for interpolation, but resistant to precise, systematic control. This work introduces a framework for explicitly mapping and manipulating latent representations using a structured Organic Learning Model (OLM) pipeline. By decomposing frames into RGB and grayscale latents, compressing temporal patterns via recurrent networks, and maintaining a memory of hashed latent states, OLM produces a stable, analyzable latent space.

We show that (1) individual object contributions can be isolated by comparing object-present vs blank-canvas latents, (2) latent differences form measurable and reusable “object vectors,” and (3) additive and subtractive operations in latent space yield predictable, controllable modifications in decoded images. Crucially, this approach goes beyond color-channel manipulation, demonstrating the potential for disentangling higher-level features such as objects, textures, and transformations. Our findings su...

Introduction Generative models such as VAEs and diffusion architectures encode inputs into high-dimensional latent spaces. While these spaces are known to contain semantic information, their internal structure is poorly understood, limiting the precision of latent manipulations. Prior work (β-VAE, InfoGAN, CLIP-guided diffusion) has explored disentanglement and semantic steering, but often in toy settings or with heuristic guidance.

We argue that a systematic, physics-inspired approach to mapping latent vectors is both possible and necessary. Using our OLM pipeline, we demonstrate controlled experiments that isolate, measure, and recombine latent contributions, moving toward a true algebra of latent space.

Related Work - Disentangled Representations: β-VAE, FactorVAE, InfoGAN – aimed at separating factors like rotation, color, shape. Mostly limited to simple datasets. - Latent Editing: GAN latent walks, StyleGAN directions, prompt interpolation in diffusion. These offer steering but not rigorous mapping. - Structured Latents: Efforts in multimodal learning and temporal VAEs suggest stability benefits from structured encodings.

Our work differs by treating latents as measurable, composable entities. Instead of steering or clustering, we map, subtract, and recombine vectors to test linearity, interaction, and entanglement.

Methodology OLM Pipeline 1. Frame Decomposition: Each frame yields RGB + BW latents via a Stable Diffusion VAE. 2. Pattern Extraction: Latents pass through a frozen Pattern LSTM to capture temporal structure. 3. Compression: A trainable Compression LSTM condenses patterns into normalized vectors. 4. Central Learning: A Central LSTM predicts next-step latents, supervised by compressed representations and novelty-driven losses. 5. Latent Memory: A hashing system stores recurring states for retrieval and comparison.

Latent Mapping Procedure - Blank-Canvas Baseline: Encode an empty frame. - Object Injection: Add a single object (e.g., a red square), encode, and compute $\Delta z = z(\text{object}) - z(\text{blank})$. - Transformation Tests: Repeat across translations, scales, and rotations to evaluate stability of Δz. - Combination: Encode multi-object scenes and compare to linear sums of object vectors. - Entanglement Analysis: Measure how object vectors interact when combined, and whether subtraction cleanly removes objects without damaging context.

Results - Channel Experiments: Red-channel subtraction demonstrated visible removal of red from predictions when α and penalty weights were tuned, proving latents can be steered in interpretable ways. - Additivity: Preliminary tests suggest partial linearity: combined-object latents approximate sums of individual object vectors but show entanglement when objects overlap. - Stability Across Transformations: Object deltas remain consistent under position/scale changes, indicating separable identity vs transform factors. - Controlled Manipulation: Latent algebra allowed both targeted suppression (removing features) and augmentation (adding features), setting the stage for higher-level manipulation.

Discussion Our results suggest that latent spaces are not arbitrary: they are structured enough to support algebraic manipulation when processed through a pipeline that stabilizes and normalizes representations. Entanglement remains a challenge — objects interact in ways that complicate clean additivity — but systematic mapping provides tools for measuring and potentially resolving these overlaps.

This establishes a precedent for latent dictionaries: reusable sets of object and transformation vectors that can be added, subtracted, and composed to engineer new scenes.

Contributions 1. A framework (OLM) that stabilizes latents for analysis and manipulation. 2. A systematic method for extracting, testing, and recombining “object vectors.” 3. Experimental evidence that latent algebra produces consistent, controllable image edits. 4. Demonstration that disentanglement and entanglement can be directly measured through controlled object-composition experiments.

Future Work - Extend beyond color/object presence into textures, lighting, and higher-level semantics. - Explore automated discovery of latent basis vectors using contrastive and orthogonalization methods. - Build a “latent dictionary” enabling reusable object and transformation controls. - Apply this methodology to dynamic video prediction tasks, testing if object latents remain stable under temporal evolution.

Conclusion This work shows that latent space is not only navigable but map-able. By treating latent vectors as measurable entities and subjecting them to controlled experiments, we can begin to understand and manipulate the algebra of representation itself. The OLM framework provides both the structure and stability needed to push beyond interpolation into true latent engineering.