

Novelty Assessment Report

Paper: Navigating the Latent Space Dynamics of Neural Models

PDF URL: <https://openreview.net/pdf?id=Zunww3FHPU>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-01

Abstract

Neural networks transform high-dimensional data into compact, structured representations, often modeled as elements of a lower dimensional latent space. In this paper, we present an alternative interpretation of neural models as dynamical systems acting on the latent manifold. Specifically, we show that autoencoder models implicitly define a latent vector field on the manifold, derived by iteratively applying the encoding-decoding map, without any additional training. We observe that standard training procedures introduce inductive biases that lead to the emergence of attractor points within this vector field. Drawing on this insight, we propose to leverage the vector field as a representation for the network, providing a novel tool to analyze the properties of the model and the data. This representation enables to: (i) analyze the generalization and memorization regimes of neural models, even throughout training; (ii) extract prior knowledge encoded in the network's parameters from the attractors, without requiring any input data; (iii) identify out-of-distribution samples from their trajectories in the vector field. We further validate our approach on vision foundation models, showcasing the applicability and effectiveness of our method in real-world scenarios.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Analyzing Latent Vector Field Dynamics in Autoencoder Neural Networks**

A total of **50 papers** were analyzed and organized into a taxonomy with **24 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Latent Space Dynamics Modeling and Prediction**
- **Latent Space Structure and Representation Learning**
- **Generative Modeling in Latent Space**
- **Applications of Latent Dynamics Analysis**
- **Theoretical Foundations and Analysis Methods**

Complete Taxonomy Tree

- Analyzing Latent Vector Field Dynamics in Autoencoder Neural Networks Survey Taxonomy
- Latent Space Dynamics Modeling and Prediction
 - Physics-Informed Latent Dynamics
 - PDE-Constrained Latent Evolution (4 papers)
 - [6] Learning physics constrained dynamics using autoencoders (TY Yang, 2022) [View paper](#)
 - [8] mLaSDI: Multi-stage latent space dynamics identification (Anderson, 2025) [View paper](#)
 - [9] Dynamically meaningful latent representations of dynamical systems (Imran Nasim, 2024) [View paper](#)
 - [42] Latent space dynamics learning for stiff collisional-radiative models (Xuping Xie, 2024) [View paper](#)
 - Koopman Operator and Linearization Methods (2 papers)
 - [2] MoriâZwanzig latent space Koopman closure for nonlinear autoencoder (Priyam Gupta, 2025) [View paper](#)
 - [7] From latent dynamics to meaningful representations (Dedi Wang, 2024) [View paper](#)
 - Neural Contractive Systems (1 papers)
 - [17] Neural contractive dynamical systems (Beik-Mohammadi, 2024) [View paper](#)
 - Data-Driven Temporal Latent Dynamics
 - Recurrent and Sequential Latent Models (7 papers)
 - [1] Accelerating phase-field simulation of coupled microstructural evolution using autoencoder-based recurrent neural networks (Aidan Gesch, 2025) [View paper](#)
 - [5] Î²-Variational autoencoders and transformers for reduced-order modelling of fluid flows (A. Solera-Rico, 2024) [View paper](#)
 - [10] Predicting turbulent dynamics with the convolutional autoencoder echo state network (Alberto Racca, 2023) [View paper](#)
 - [11] Towards latent space evolution of spatiotemporal dynamics of six-dimensional phase space of charged particle beams (Rautela, 2024) [View paper](#)
 - [13] Dynamical variational autoencoders: A comprehensive review (Girin, 2022) [View paper](#)
 - [48] Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes (K. Hasegawa, 2020) [View paper](#)
 - [49] Latent space physics: Towards learning the temporal evolution of fluid flow (Steffen Wiewel, 2019) [View paper](#)
 - State-Space Latent Dynamics (2 papers)
 - [19] Deep latent state space models for time-series generation (Zhou, 2023) [View paper](#)
 - [43] Flow-Based Variational Sequence Autoencoder (Jen-Tzung Chien, 2022) [View paper](#)
 - Latent Dynamics for Specific Physical Systems (2 papers)
 - [22] Flow time history representation and reconstruction based on machine learning (Qingliang Zhan, 2023) [View paper](#)

- [26] Dynamics of a data-driven low-dimensional model of Rayleigh-Benard convection (Chen Qi-wei, 2025) [View paper](#)
- Latent Space Structure and Representation Learning
 - Latent Space Geometry and Manifold Learning (4 papers)
 - [20] A novel method of low-dimensional representation for temporal behavior of flow fields using deep autoencoder (Noriyasu Omata, 2019) [View paper](#)
 - [27] Autoencoders for discovering manifold dimension and coordinates in data from complex dynamical systems (Kevin Zeng, 2024) [View paper](#)
 - [33] The most robust representations of flow trajectories are Lagrangian coherent structures (Theodore MacMillan, 2021) [View paper](#)
 - [46] Variational autoencoder with learned latent structure (Marissa Connor, 2021) [View paper](#)
 - Latent Space Exploration and Intervention (3 papers)
 - [15] Synergizing Attribute-Guided Latent Space Exploration (AGLSE) with Classical Molecular Simulations to Design Potent Peptide Inhibitors to Abrogate SARS-CoV-2 Host Cell Entry (Farhan Ullah, 2025) [View paper](#)
 - [18] Exploring the latent space of autoencoders with interventional assays (Leeb, 2022) [View paper](#)
 - [28] Grammarization-based grasping with deep multi-autoencoder latent space exploration by reinforcement learning agent (Askianakis, 2024) [View paper](#)
 - Specialized Latent Representations
 - Continuous and Parametric Latent Representations (2 papers)
 - [3] CLR-Wire: Towards Continuous Latent Representations for 3D Curve Wireframe Generation (Ma Xueqi, 2025) [View paper](#)
 - [35] Encoding and exploring latent design space of optimal material structures via a VAE-LSTM model (Andrew J. Lew, 2021) [View paper](#)
 - Vector Quantized Latent Spaces (1 papers)
 - [29] Vector quantization variational autoencoder for turbulent flow images (Daryl Fung, 2025) [View paper](#)
 - Hierarchical and Multi-Level Latent Structures (1 papers)
 - [25] WF-VAE: Enhancing Video VAE by Wavelet-Driven Energy Flow for Latent Video Diffusion Model (Li Zongjian, 2024) [View paper](#)
 - Latent Space Alignment and Robustness (3 papers)
 - [30] Enabling dynamic 3D coherent diffraction imaging via adaptive latent space tuning of generative autoencoders (Alexander Scheinker, 2024) [View paper](#)
 - [37] Adaptive autoencoder latent space tuning for more robust machine learning beyond the training set for six-dimensional phase space diagnostics of a time-varying $\hat{\alpha}$ (A Scheinker, 2023) [View paper](#)
 - [44] Using Optimal Transport Aligned Latent Embeddings for Separated Flow Analysis (Yeh, 2025) [View paper](#)
- Generative Modeling in Latent Space
 - Flow-Based and Diffusion Latent Generation (2 papers)
 - [12] Flow matching in latent space (Dao-quan, 2023) [View paper](#)
 - [41] EchoFlow: A Foundation Model for Cardiac Ultrasound Image and Video Generation (Reynaud, 2025) [View paper](#)
 - Variational Latent Generation (2 papers)
 - [32] Developing an Explainable Variational Autoencoder (VAE) Framework for Accurate Representation of Local Circulation in Taiwan (Min-Ken Hsieh, 2024) [View paper](#)
 - [38] Generalized Molecular Latent Representation via Graph Latent Diffusion Autoencoder (Daiki Koge, 2025) [View paper](#)
 - Latent Interpolation and Synthesis (1 papers)
 - [50] Generating In-Between Images Through Learned Latent Space Representation Using Variational Autoencoders (Paulino Cristovao, 2020) [View paper](#)
- Applications of Latent Dynamics Analysis
 - Anomaly and Novelty Detection (2 papers)
 - [4] Latent space autoregression for novelty detection (Davide Abati, 2019) [View paper](#)
 - [16] Enhancing Anomaly Detection Through Latent Space Manipulation in Autoencoders: A Comparative Analysis (Tomasz Walczyna, 2024) [View paper](#)
 - Control and Optimization via Latent Dynamics (1 papers)
 - [23] Big Data-Driven Control of Nonlinear Processes Through Dynamic Latent Variables Using an Autoencoder (Jun Wen Tang, 2025) [View paper](#)
 - Reduced-Order Modeling for Engineering (3 papers)
 - [21] Latent space dynamics for reduced deformable simulation (Lawson Fulton, 2019) [View paper](#)
 - [39] Reduced order modeling using advection-aware autoencoders (Sourav Dutta, 2022) [View paper](#)
 - [45] Nonlinear mode decomposition via physics-assimilated convolutional autoencoder for unsteady flows over an airfoil (Bo Zhang, 2023) [View paper](#)
 - Biological and Neuroscience Applications (2 papers)
 - [14] Reconstructing developmental trajectories using latent dynamical systems and time-resolved transcriptomics (R. Maizels, 2024) [View paper](#)
 - [31] Reconstructing feedback representations in the ventral visual pathway with a generative adversarial autoencoder (Haider Al-Tahan, 2021) [View paper](#)
 - Autonomous Systems and Robotics (1 papers)
 - [36] Carnet: A dynamic autoencoder for learning latent dynamics in autonomous driving tasks (Pak, 2022) [View paper](#)
- Theoretical Foundations and Analysis Methods
 - Latent Vector Field Theory and Attractors ★ (1 papers)
 - [0] Navigating the Latent Space Dynamics of Neural Models (Anon et al., 2026) [View paper](#)
 - Comparative Studies and Frameworks (1 papers)
 - [47] Autoencoder embeddings in dynamical systems: A comparison of current frameworks (Quinn, 2024) [View paper](#)
 - Incremental and Continual Latent Learning (2 papers)
 - [34] Preserving Seasonal and Trend Information: A Variational Autoencoder-Latent Space Arithmetic Based Approach for Non-Stationary Learning (Hassan Wasswa, 2024) [View paper](#)
 - [40] Learning additional latent representations using variational autoencoders (Chong Shen, 2023) [View paper](#)
 - Exploration of Latent Dynamics Representations (1 papers)
 - [24] Exploration and prediction of fluid dynamical systems using auto-encoder technology (Lionel, 2020) [View paper](#)

Narrative

Core task: analyzing latent vector field dynamics in autoencoder neural networks. The field organizes around five main branches that reflect complementary perspectives on how autoencoders compress and evolve high-dimensional data. Latent Space Dynamics Modeling and Prediction focuses on forecasting temporal evolution within learned representations, often coupling reduced-order models with neural architectures. Latent Space Structure and Representation Learning examines the geometric and topological properties that emerge during encoding, asking how disentanglement, manifold structure, and interpretability arise. Generative Modeling in Latent Space emphasizes sampling and synthesis, leveraging variational or flow-based frameworks to produce novel instances. Applications of Latent Dynamics Analysis demonstrates these methods in domains ranging from fluid mechanics and materials science to biological trajectory reconstruction. Finally, Theoretical Foundations and Analysis Methods provides the mathematical underpinnings—attractor theory, stability analysis, and operator-theoretic perspectives—that justify why latent dynamics can faithfully capture complex system behavior.

Several active lines of work reveal key trade-offs between physical interpretability and expressive power. Physics-informed approaches such as Physics Constrained Autoencoders[6] and mLaSDI[8] embed known governing equations directly into the latent space, ensuring that learned dynamics respect conservation laws or symmetries, while purely data-driven methods like Dynamically Meaningful Latent[9] and Latent Space Evolution[11] prioritize flexibility and scalability at the cost of harder-to-interpret vector fields. Navigating Latent Space[0] sits within the Theoretical Foundations branch, specifically addressing latent vector field theory and attractors. Its emphasis on rigorous characterization of fixed points and flow topology aligns it closely with works like Mori-Zwanzig Koopman[2] and Neural Contractive Systems[17], which also seek formal guarantees on latent dynamics. Compared to application-focused studies such as Fluid Dynamics Autoencoder[24] or Data-Driven Rayleigh-Benard[26], Navigating Latent Space[0] offers a more foundational lens, exploring how attractor structure in latent space relates to the original high-dimensional system's long-term behavior.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

The original leaf focuses on the theoretical analysis of vector fields, attractors, and implicit dynamical systems within autoencoder latent spaces, emphasizing mathematical foundations. The sibling subtopics address complementary aspects: comparative frameworks for surveying methods, representational properties of latent dynamics, and temporal adaptation mechanisms. While the original leaf is theory-driven, siblings span empirical comparisons, interpretive analysis, and continual learning applications.

Similarities: - All subtopics operate within the autoencoder latent space context and concern dynamics or temporal evolution - Each excludes overlapping concerns through explicit scope boundaries (e.g., all distinguish from empirical method development or static representation learning) - Shared interest in understanding latent space behavior, whether through theoretical models, comparative analysis, or adaptive mechanisms

Differences: - Original leaf emphasizes mathematical theory (vector fields, attractors, implicit systems) while siblings focus on empirical frameworks, representational interpretation, or incremental learning - Comparative Studies examines multiple methods horizontally; Exploration analyzes representational semantics; Incremental Learning addresses non-stationarity; original leaf develops foundational dynamical theory - Original leaf excludes empirical method development, while Exploration and Incremental Learning may involve method proposals for their specific scopes - Temporal scope differs: original leaf studies intrinsic dynamics, Incremental Learning handles external temporal changes in data distribution

Suggested Search Directions: - Connections between theoretical attractor analysis and empirical continual learning stability - How vector field theory informs comparative evaluation metrics for latent dynamics methods - Representational properties that emerge from specific attractor configurations

Sibling Subtopics

- **Comparative Studies and Frameworks** (leaves: 1, papers: 1)
 - Scope: Surveys, reviews, or comparative analyses of autoencoder architectures and latent dynamics methods.
 - Exclude: Excludes novel method proposals; see Latent Space Dynamics Modeling and Prediction.
- **Exploration of Latent Dynamics Representations** (leaves: 1, papers: 1)
 - Scope: Studies analyzing how latent dynamics serve as representations for model properties or data characteristics.
 - Exclude: Excludes theoretical vector field analysis; see Latent Vector Field Theory and Attractors.
- **Incremental and Continual Latent Learning** (leaves: 1, papers: 2)
 - Scope: Methods for learning additional latent representations or adapting to non-stationary data over time.
 - Exclude: Excludes static representation learning; see Latent Space Structure and Representation Learning.

Contributions Analysis

Overall novelty summary. The paper proposes interpreting autoencoders as dynamical systems by defining a latent vector field through iterative encoding-decoding, identifying attractor points that emerge from standard training. It resides in the 'Latent Vector Field Theory and Attractors' leaf under 'Theoretical Foundations and Analysis Methods', which currently contains only this paper among 50 total papers in the taxonomy. This isolation suggests the work occupies a relatively sparse theoretical niche, focusing on formal characterization of implicit dynamics rather than method development or domain applications.

The taxonomy reveals substantial activity in neighboring branches: 'Latent Space Dynamics Modeling and Prediction' contains 19 papers across physics-informed and data-driven temporal modeling, while 'Latent Space Structure and Representation Learning' includes 13 papers on geometry and manifold discovery. The paper's theoretical focus on attractor dynamics connects it to 'Neural Contractive Systems' and 'Koopman Operator' methods within physics-informed dynamics, yet diverges by analyzing implicit vector fields in standard autoencoders rather than designing architectures with explicit stability constraints. Its position bridges foundational theory and the broader dynamics modeling literature.

Among 22 candidates examined, the contribution on latent vector field representation found no refuting prior work across 10 candidates, suggesting novelty in framing autoencoders as implicit dynamical systems. However, the memorization-generalization connection via attractors encountered 1 refutable candidate among 10 examined, indicating some overlap with existing analyses of training dynamics. The data-free probing contribution examined only 2 candidates with no refutations, though the limited search scope leaves open the possibility of undetected prior work in foundation model analysis or noise-based probing techniques.

Based on top-22 semantic matches, the vector field interpretation and attractor-based analysis appear relatively novel within the examined scope, particularly the formal treatment of implicit dynamics in standard autoencoders. The memorization-generalization link shows partial overlap with prior training dynamics research, while the foundation model probing contribution remains underexplored in this limited search. The sparse population of the theoretical attractors leaf and the paper's bridging position between theory and applications suggest it addresses a gap, though exhaustive coverage of related dynamical systems theory or representation learning literature cannot be claimed.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Latent vector field representation of autoencoders

Description: The authors introduce a novel interpretation of autoencoder models as dynamical systems that implicitly define a latent vector field through iterative application of the encoding-decoding map. This vector field arises naturally without requiring additional training and provides a new tool for analyzing model and data properties.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. mLaSDI: Multi-stage latent space dynamics identification

URL: [View paper](#)

Brief Assessment

mLaSDI[8] focuses on multi-stage training of autoencoders for reduced-order modeling of PDEs with explicit dynamics identification (SINDy), not on interpreting autoencoders as dynamical systems with implicit latent vector fields arising from iterative encoding-decoding.

2. The autoencoding variational autoencoder

URL: [View paper](#)

Brief Assessment

Autoencoding VAE[62] focuses on encoder-decoder consistency through iterative encoding-decoding to achieve adversarial robustness, not on analyzing latent vector fields as dynamical systems or studying their attractors and trajectories as tools for model analysis.

3. Inverse Problem Sampling in Latent Space Using Sequential Monte Carlo

URL: [View paper](#)

Brief Assessment

Inverse Latent Sampling[63] focuses on inverse problem solving using SMC sampling in the latent space of diffusion models, not on analyzing autoencoders as dynamical systems with latent vector fields arising from iterative encoding-decoding.

4. Towards latent space evolution of spatiotemporal dynamics of six-dimensional phase space of charged particle beams

URL: [View paper](#)

Brief Assessment

Latent Space Evolution[11] uses autoencoders (specifically CVAEs) combined with LSTMs for temporal forecasting of particle beam dynamics, not for defining implicit latent vector fields through iterative encoding-decoding maps as in the original paper.

5. Hyperspectral band selection with iterative graph autoencoder

URL: [View paper](#)

Brief Assessment

Iterative Graph Autoencoder[64] focuses on hyperspectral band selection using graph autoencoders with iterative refinement for structure representation. It does not address latent vector fields arising from iterative encoding-decoding maps in general autoencoder architectures.

6. Decoding Vocal Articulations from Acoustic Latent Representations

URL: [View paper](#)

Brief Assessment

Decoding Vocal Articulations[67] focuses on acoustic-to-articulatory inversion using variational autoencoders for voice synthesis parameter prediction, not on interpreting autoencoders as dynamical systems with latent vector fields through iterative encoding-decoding.

7. REED-VAE: RE-Encode Decode Training for Iterative Image Editing with Diffusion Models

URL: [View paper](#)

Brief Assessment

REED-VAE[65] focuses on training variational autoencoders for iterative image editing quality preservation, not on interpreting autoencoders as dynamical systems with latent vector fields arising from iterative encoding-decoding maps.

8. Collaborative Filtering Algorithm Based on Deep Denoising Auto-Encoder and Attention Mechanism

URL: [View paper](#)

Brief Assessment

Deep Denoising Collaborative[61] focuses on collaborative filtering using denoising autoencoders for recommendation systems, not on interpreting autoencoders as dynamical systems with latent vector fields through iterative encoding-decoding.

9. Latent space dynamics learning for stiff collisional-radiative models

URL: [View paper](#)

Brief Assessment

Latent Collisional-Radiative[42] focuses on learning latent dynamics for collisional-radiative models in plasma physics using flow maps, not on interpreting autoencoders as dynamical systems with implicit vector fields arising from iterative encoding-decoding.

10. AROMA: Preserving spatial structure for latent PDE modeling with local neural fields

URL: [View paper](#)

Brief Assessment

AROMA[66] focuses on latent representations for PDE modeling using local neural fields with encoder-decoder architectures, but does not discuss latent vector fields arising from iterative encoding-decoding maps or their use as dynamical systems for analyzing model properties.

Contribution 2: Connection between attractors and memorization-generalization regimes

Description: The work demonstrates that attractors in the latent vector field encode whether a model is in a memorization or generalization regime. The authors show empirically how these attractors evolve throughout the training process, providing insights into the learning dynamics of neural networks.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Memorization to generalization: Emergence of diffusion models from associative memory

URL: [View paper](#)

Brief Assessment

Diffusion from Memory[53] demonstrates that attractors in energy landscapes encode whether a diffusion model is in memorization or generalization regimes, and empirically shows how these attractors evolve throughout training. The paper provides a comprehensive framework connecting attractor dynamics to the memorization-generalization transition, including theoretical predictions and empirical validation across multiple datasets. The work explicitly characterizes three distinct phases (memorization, spurious, generalization) through attractor behavior and provides metrics to detect these phases during training.

Evidence

Evidence 1 - **Rationale:** Both papers establish that attractors indicate memorization versus generalization regimes. The candidate explicitly describes how basins of attraction around samples characterize memorization, and how new attractor states emerge during generalization. - **Original:** we empirically connect attractors to the network's memorization and generalization regimes, showing how they evolve during training. - **Candidate:** in the small data regime the diffusion model exhibits a strong memorization phase, where the network creates distinct basins of attraction around each sample in the training set, akin to the hopfield model below the critical memory load. in the large data regime, a different phase appears where an i...

Evidence 2 - **Rationale:** Both papers empirically demonstrate the memorization-to-generalization transition and show how attractors/samples evolve during this transition, providing similar empirical evidence for the phenomenon. - **Original:** in figure 3b we show the transition from memorization, occurring at the first epochs of training, to generalization, by plotting the memorization coefficient and the test error across training, observing a trade-off between the two. - **Candidate:** the results in fig. (5) clearly demonstrate the transition from memorization to generalization as the dataset size increases. meanwhile, the collected samples also show distinct characteristics in each of the considered phases

Evidence 3 - **Rationale:** Both papers characterize attractors as capturing the memorization-generalization interplay. The candidate provides additional detail about spurious states at the transition boundary, demonstrating prior work on attractor behavior during this transition. - **Original:** attractors capture the interplay between generalization and memorization of neural models, which corresponds to the trade-off between the reconstruction performance and regularization term of the ae model. - **Candidate:** spurious states appear at the boundary of this transition and correspond to emergent attractor states, which are absent in the training set, but, at the same time, have distinct basins of attraction around them.

2. Pseudo-likelihood produces associative memories able to generalize, even for asymmetric couplings

URL: [View paper](#)

Brief Assessment

Pseudo-Likelihood Associative[57] focuses on energy-based models trained via pseudo-likelihood maximization, studying how attractors in these models relate to memorization versus generalization. The original paper examines attractors in autoencoder latent vector fields during training. These are fundamentally different model architectures and training objectives, so this candidate does not challenge the novelty of the original contribution.

3. Self-orthogonalizing attractor neural networks emerging from the free energy principle

URL: [View paper](#)

Brief Assessment

Self-Orthogonalizing Attractor[52] focuses on attractor networks derived from the free energy principle in neuroscience, not on analyzing memorization versus generalization regimes during neural network training as studied in the original paper.

4. Attractor Regimes of Boolean Recurrent Neural Networks subject to STDP and Global Plasticity

URL: [View paper](#)

Brief Assessment

Boolean RNN Attractors[59] focuses on attractor dynamics in Boolean recurrent neural networks for memory mechanisms, not on how attractors indicate memorization versus generalization regimes during neural network training.

5. State-denoised recurrent neural networks

URL: [View paper](#)

Brief Assessment

State-Denoised RNN[58] focuses on using attractor dynamics to denoise hidden states in recurrent networks for improved generalization on sequence tasks. It does not investigate how attractors encode memorization versus generalization regimes during training.

6. Why do recurrent neural networks suddenly learn? Bifurcation mechanisms in neuro-inspired short-term memory tasks

URL: [View paper](#)

Brief Assessment

RNN Bifurcation Mechanisms[56] focuses on bifurcations during RNN training for short-term memory tasks, not on how attractors encode memorization versus generalization regimes in autoencoders. The candidate studies attractor formation as a mechanism for task learning, while the original work analyzes attractors as indicators of model generalization state.

7. Reinforcing Neural Network Stability with Attractor Dynamics

URL: [View paper](#)

Brief Assessment

Reinforcing Network Stability[60] focuses on using attractor dynamics to improve stability and generalization in ResNet-like architectures through a training-time module (RMAN). It does not examine how attractors evolve during training to indicate memorization versus generalization regimes, which is the core novelty claim of the original paper.

8. Training neural networks with structured noise improves classification and generalization

URL: [View paper](#)

Brief Assessment

Structured Noise Training[55] focuses on noise-injection training algorithms for recurrent networks and memory retrieval, not on analyzing how attractors in latent vector fields indicate memorization versus generalization regimes during neural network training.

9. Analytical Methods for Continuous Attractor Neural Networks

URL: [View paper](#)

Brief Assessment

Continuous Attractor Networks[51] focuses on analytical methods for continuous attractor neural networks in spatial memory contexts (place cells, hippocampus), not on the memorization-generalization dynamics during neural network training that the original paper investigates.

10. Line Attractor Dynamics for Latent Space Regularization in Deep Neural Networks

URL: [View paper](#)

Brief Assessment

Line Attractor Dynamics[54] focuses on polynomial latent space regularization in deep neural networks for robustness and generalization, not on analyzing how attractors in latent vector fields indicate memorization versus generalization regimes during training as in the original paper.

Contribution 3: Data-free probing of foundation models via noise-derived attractors

Description: The authors propose a method to extract knowledge encoded in pretrained foundation models without requiring any input data. By computing attractors from Gaussian noise initialization, they can recover semantic information stored in the network weights, enabling black-box analysis of model representations.

This contribution was assessed against **2 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Concept Attractors in LLMs and their Applications

URL: [View paper](#)

Brief Assessment

Concept Attractors LLMs[69] focuses on LLMs and uses attractors for downstream tasks like translation and guardrailing, not on data-free probing of foundation models' encoded knowledge through noise-derived attractors as in the original paper's autoencoder framework.

2. Can Generative Geospatial Diffusion Models Excel as Discriminative Geospatial Foundation Models?

URL: [View paper](#)

Brief Assessment

Generative Geospatial Diffusion[68] focuses on transforming diffusion-based generative models into discriminative tools for remote sensing tasks, not on data-free probing of foundation models using noise-derived attractors.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Navigating the Latent Space Dynamics of Neural Models [View paper](#)
- [1] Accelerating phase-field simulation of coupled microstructural evolution using autoencoder-based recurrent neural networks [View paper](#)
- [2] Moriá[Zwanzig latent space Koopman closure for nonlinear autoencoder [View paper](#)
- [3] CLR-Wire: Towards Continuous Latent Representations for 3D Curve Wireframe Generation [View paper](#)
- [4] Latent space autoregression for novelty detection [View paper](#)
- [5] \hat{I}^2 -Variational autoencoders and transformers for reduced-order modelling of fluid flows [View paper](#)
- [6] Learning physics constrained dynamics using autoencoders [View paper](#)
- [7] From latent dynamics to meaningful representations [View paper](#)
- [8] mLaSDI: Multi-stage latent space dynamics identification [View paper](#)
- [9] Dynamically meaningful latent representations of dynamical systems [View paper](#)
- [10] Predicting turbulent dynamics with the convolutional autoencoder echo state network [View paper](#)
- [11] Towards latent space evolution of spatiotemporal dynamics of six-dimensional phase space of charged particle beams [View paper](#)
- [12] Flow matching in latent space [View paper](#)
- [13] Dynamical variational autoencoders: A comprehensive review [View paper](#)
- [14] Reconstructing developmental trajectories using latent dynamical systems and time-resolved transcriptomics [View paper](#)
- [15] Synergizing Attribute-Guided Latent Space Exploration (AGLSE) with Classical Molecular Simulations to Design Potent Peptide Inhibitors to Abrogate SARS-CoV-2 Host Cell Entry [View paper](#)
- [16] Enhancing Anomaly Detection Through Latent Space Manipulation in Autoencoders: A Comparative Analysis [View paper](#)
- [17] Neural contractive dynamical systems [View paper](#)
- [18] Exploring the latent space of autoencoders with interventional assays [View paper](#)
- [19] Deep latent state space models for time-series generation [View paper](#)
- [20] A novel method of low-dimensional representation for temporal behavior of flow fields using deep autoencoder [View paper](#)
- [21] Latent space dynamics for reduced deformable simulation [View paper](#)
- [22] Flow time history representation and reconstruction based on machine learning [View paper](#)
- [23] Big Data-Driven Control of Nonlinear Processes Through Dynamic Latent Variables Using an Autoencoder [View paper](#)
- [24] Exploration and prediction of fluid dynamical systems using auto-encoder technology [View paper](#)
- [25] WF-VAE: Enhancing Video VAE by Wavelet-Driven Energy Flow for Latent Video Diffusion Model [View paper](#)
- [26] Dynamics of a data-driven low-dimensional model of Rayleigh-Benard convection [View paper](#)
- [27] Autoencoders for discovering manifold dimension and coordinates in data from complex dynamical systems [View paper](#)
- [28] Grammarization-based grasping with deep multi-autoencoder latent space exploration by reinforcement learning agent [View paper](#)
- [29] Vector quantization variational autoencoder for turbulent flow images [View paper](#)
- [30] Enabling dynamic 3D coherent diffraction imaging via adaptive latent space tuning of generative autoencoders [View paper](#)
- [31] Reconstructing feedback representations in the ventral visual pathway with a generative adversarial autoencoder [View paper](#)

- [32] Developing an Explainable Variational Autoencoder (VAE) Framework for Accurate Representation of Local Circulation in Taiwan [View paper](#)
- [33] The most robust representations of flow trajectories are Lagrangian coherent structures [View paper](#)
- [34] Preserving Seasonal and Trend Information: A Variational Autoencoder-Latent Space Arithmetic Based Approach for Non-Stationary Learning [View paper](#)
- [35] Encoding and exploring latent design space of optimal material structures via a VAE-LSTM model [View paper](#)
- [36] Carnet: A dynamic autoencoder for learning latent dynamics in autonomous driving tasks [View paper](#)
- [37] Adaptive autoencoder latent space tuning for more robust machine learning beyond the training set for six-dimensional phase space diagnostics of a time-varying $\hat{\alpha}$; [View paper](#)
- [38] Generalized Molecular Latent Representation via Graph Latent Diffusion Autoencoder [View paper](#)
- [39] Reduced order modeling using advection-aware autoencoders [View paper](#)
- [40] Learning additional latent representations using variational autoencoders [View paper](#)
- [41] EchoFlow: A Foundation Model for Cardiac Ultrasound Image and Video Generation [View paper](#)
- [42] Latent space dynamics learning for stiff collisional-radiative models [View paper](#)
- [43] Flow-Based Variational Sequence Autoencoder [View paper](#)
- [44] Using Optimal Transport Aligned Latent Embeddings for Separated Flow Analysis [View paper](#)
- [45] Nonlinear mode decomposition via physics-assimilated convolutional autoencoder for unsteady flows over an airfoil [View paper](#)
- [46] Variational autoencoder with learned latent structure [View paper](#)
- [47] Autoencoder embeddings in dynamical systems: A comparison of current frameworks [View paper](#)
- [48] Machine-learning-based reduced-order modeling for unsteady flows around bluff bodies of various shapes [View paper](#)
- [49] Latent space physics: Towards learning the temporal evolution of fluid flow [View paper](#)
- [50] Generating In-Between Images Through Learned Latent Space Representation Using Variational Autoencoders [View paper](#)
- [51] Analytical Methods for Continuous Attractor Neural Networks [View paper](#)
- [52] Self-orthogonalizing attractor neural networks emerging from the free energy principle [View paper](#)
- [53] Memorization to generalization: Emergence of diffusion models from associative memory [View paper](#)
- [54] Line Attractor Dynamics for Latent Space Regularization in Deep Neural Networks [View paper](#)
- [55] Training neural networks with structured noise improves classification and generalization [View paper](#)
- [56] Why do recurrent neural networks suddenly learn? Bifurcation mechanisms in neuro-inspired short-term memory tasks [View paper](#)
- [57] Pseudo-likelihood produces associative memories able to generalize, even for asymmetric couplings [View paper](#)
- [58] State-denoised recurrent neural networks [View paper](#)
- [59] Attractor Regimes of Boolean Recurrent Neural Networks subject to STDP and Global Plasticity [View paper](#)
- [60] Reinforcing Neural Network Stability with Attractor Dynamics [View paper](#)
- [61] Collaborative Filtering Algorithm Based on Deep Denoising Auto-Encoder and Attention Mechanism [View paper](#)
- [62] The autoencoding variational autoencoder [View paper](#)
- [63] Inverse Problem Sampling in Latent Space Using Sequential Monte Carlo [View paper](#)
- [64] Hyperspectral band selection with iterative graph autoencoder [View paper](#)
- [65] REED-VAE: RE-Encode Decode Training for Iterative Image Editing with Diffusion Models [View paper](#)
- [66] AROMA: Preserving spatial structure for latent PDE modeling with local neural fields [View paper](#)
- [67] Decoding Vocal Articulations from Acoustic Latent Representations [View paper](#)
- [68] Can Generative Geospatial Diffusion Models Excel as Discriminative Geospatial Foundation Models? [View paper](#)
- [69] Concept Attractors in LLMs and their Applications [View paper](#)