Novelty Assessment Report

Paper: Diffusion Transformers with Representation Autoencoders

PDF URL: https://openreview.net/pdf?id=0u1LigJaab

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-05

Abstract

Latent generative modeling has become the standard strategy for Diffusion Transformers (DiTs), but the autoencoder has barely evolved. Most DiTs still use the legacy VAE encoder, which introduces several limitations: large convolutional backbones that compromise architectural simplicity, low-dimensional latent spaces that restrict information capacity, and weak representations resulting from purely reconstruction-based training. In this work, we investigate replacing the VAE encoder–decoder with pretrained representation encoders (e.g., DINO, SigLIP, MAE) combined with trained decoders, forming what we call \emph{Representation Autoencoders} (RAEs). These models provide both high-quality reconstructions and semantically rich latent spaces, while allowing for a scalable transformer-based architecture. A key challenge arises in enabling diffusion transformers to operate effectively within these high-dimensional representations. We analyze the sources of this difficulty, propose theoretically motivated solutions, and validate them empirically. Our approach achieves faster convergence without auxiliary representation alignment losses. Using a DiT variant with a lightweight wide DDT-head, we demonstrate state-of-the-art image generation performance, reaching FIDs of 1.18 @256 resolution and 1.13 @512 on ImageNet.

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: **Latent Diffusion Modeling with Representation Autoencoders**

A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

Taxonomy Overview

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

- **Autoencoder Architecture and Latent Space Design**
- **Representation Learning and Semantic Encoding**
- **Generation Applications Across Domains**
- **Specialized Applications and Analysis**

Complete Taxonomy Tree

- Latent Diffusion Modeling with Representation Autoencoders Survey Taxonomy
- Autoencoder Architecture and Latent Space Design
  - High-Compression and Efficient Autoencoders (3 papers)
  - [8] Improving the diffusability of autoencoders (Skorokhodov, 2025) View paper
  - [13] Deep compression autoencoder for efficient high-resolution diffusion models (Chen, 2024) View paper
  - [18] Dc-ae 1.5: Accelerating diffusion model convergence with structured latent space (Chen, 2025) View paper
  - Latent Space Properties and Optimization (4 papers)
  - [11] Latent Diffusion Models with Masked AutoEncoders (Lee Junho, 2025) View paper
  - [30] Litevae: Lightweight and efficient variational autoencoders for latent diffusion models (Derek Bradley, 2024) View paper
  - [47] Masked autoencoders are effective tokenizers for diffusion models (Chen, 2025) View paper
  - [49] Binary latent diffusion (Ze Wang, 2023) View paper
  - Masked Autoencoder Integration (1 papers)
  - [2] Diffusion models as masked autoencoders (Chen Wei, 2023) View paper
  - Unified End-to-End Architectures (2 papers)
  - [21] Diffusion As Self-Distillation: End-to-End Latent Diffusion In One Model (Xiyuan Wang, 2025) View paper
  - [45] DGAE: Diffusion-Guided Autoencoder for Efficient Latent Representation Learning (Liu Dongxu, 2025) View paper
  - Representation Autoencoders ★ (2 papers)
  - [0] Diffusion Transformers with Representation Autoencoders (Anon et al., 2026) View paper
  - [14] MeanFlow Transformers with Representation Autoencoders (Zheyuan Hu, 2025) View paper
- Representation Learning and Semantic Encoding
  - Diffusion-Based Representation Learning (4 papers)
  - [3] Diffusion model as representation learner (Xing-yi Yang, 2023) View paper
  - [5] Diffusion based representation learning (Mittal, 2023) View paper
  - [40] Representation learning with diffusion models (Traub, 2022) View paper
  - [46] Automated Learning of Semantic Embedding Representations for Diffusion Models (Jiang, 2025) View paper
  - Disentangled Representation Learning (3 papers)
  - [7] Factorized diffusion autoencoder for unsupervised disentangled representation learning (Wu, 2024) View paper
  - [20] Hierarchical diffusion autoencoders and disentangled image manipulation (Zeyu Lu, 2024) View paper
  - [36] DiffSDA: Unsupervised Diffusion Sequential Disentanglement Across Modalities (Zisling, 2025) View paper
  - Semantic Autoencoding and Reconstruction (1 papers)

    - [27] Diffusion autoencoders: Toward a meaningful and decodable representation (Preechakul, 2022) View paper
    - Bridge Autoencoders (1 papers)
    - [15] Diffusion Bridge AutoEncoders for Unsupervised Representation Learning (Kim Yeongmin, 2024) View paper
- Generation Applications Across Domains
    - Image and Visual Content Generation (1 papers)
    - [28] High-Resolution Image Synthesis with Latent Diffusion Models (Robin Rombach, 2021) View paper
    - 3D and Geometric Generation (4 papers)
    - [9] Lion: Latent point diffusion models for 3d shape generation (Zeng, 2022) View paper
    - [26] Latent 3d graph diffusion (Y You, 2024) View paper
    - [42] LN3DIFF++: Scalable Latent Neural Fields Diffusion for Speedy 3D Generation. (Lan, 2024) View paper
    - [43] Autodecoding latent 3d diffusion models (Ntavelis, 2023) View paper
    - Molecular and Protein Generation (5 papers)
    - [6] A latent diffusion model for protein structure generation (Fu Cong, 2024) View paper
    - [19] Geometric Latent Diffusion Models for 3D Molecule Generation (Xu, 2023) View paper
    - [22] Crystal diffusion variational autoencoder for periodic material generation (Tian Xie, 2021) View paper
    - [37] ProteinAE: Protein Diffusion Autoencoders for Structure Encoding (Li Shaoning, 2025) View paper
    - [41] GLDM: hit molecule generation with constrained graph latent diffusion model (Conghao Wang, 2024) View paper
    - Sequence and Temporal Generation (5 papers)
    - [4] Latent diffusion for language generation (Lovelace, 2023) View paper
    - [10] Latent diffusion model for dna sequence generation (Li Zehui, 2023) View paper
    - [31] Executing your commands via motion diffusion in latent space (Chen Xin, 2023) View paper
    - [38] TimeAutoDiff: Combining Autoencoder and Diffusion model for time series tabular data synthesizing (Suh, 2024) View paper
    - [44] Timeldm: Latent diffusion model for unconditional time series generation (Qian Jian, 2024) View paper
    - Multimodal Generation (2 papers)
    - [23] Dae-talker: High fidelity speech-driven talking face generation with diffusion autoencoder (Chenpeng Du, 2023) View paper
    - [50] MM-LDM: Multi-Modal Latent Diffusion Model for Sounding Video Generation (Sun, 2024) View paper
    - Tabular and Structured Data Generation (1 papers)
    - [35] Mixed-Type Tabular Data Synthesis with Score-based Diffusion in Latent Space (Zhang Hengrui, 2023) View paper
- Specialized Applications and Analysis
    - Scientific and Physical System Modeling (3 papers)
    - [1] Latent diffusion modeling of porous media informed by spatial statistics. (Pejman Tahmasebi, 2025) View paper
    - [33] Lost in Latent Space: An Empirical Study of Latent Diffusion Models for Physics Emulation (Rozet, 2025) View paper
    - [34] cDVAE: Multimodal generative conditional diffusion guided by variational autoencoder latent embedding for virtual 6D phase space diagnostics (Scheinker, 2024) View paper
    - Neural and Medical Data Modeling (2 papers)
    - [12] Latent Diffusion Autoencoders: Toward Efficient and Meaningful Unsupervised Representation Learning in Medical Imaging (Bria, 2025) View paper
    - [17] Latent diffusion for neural spiking data (Kapoor, 2024) View paper
    - Anomaly Detection and Forensics (2 papers)
    - [16] Aeroblade: Training-free detection of latent diffusion images using autoencoder reconstruction error (Jonas Ricker, 2024) View paper
    - [24] Lafite: Latent diffusion model with feature editing for unsupervised multi-class anomaly detection (Yin, 2023) View paper
    - Causal and Counterfactual Generation (1 papers)
    - [32] Causal Diffusion Autoencoders: Toward Counterfactual Generation via Diffusion Probabilistic Models (Xintao Wu, 2024) View paper
    - Aging and Temporal Transformation (1 papers)
    - [39] Pluralistic Aging Diffusion Autoencoder (Peipei Li, 2023) View paper
    - Steganography and Data Hiding (1 papers)
    - [48] RoSteALS: Robust Steganography using Autoencoder Latent Space (Tu Bui, 2023) View paper
    - Generalized Frameworks and Unified Models (2 papers)
    - [25] H-space sparse autoencoders (A Ijishakin, 2024) View paper
    - [29] Unified generation, reconstruction, and representation: generalized diffusion with adaptive latent encoding-decoding (Liu Guangyi, 2024) View paper

Narrative

Core task: latent diffusion modeling with representation autoencoders. This field centers on learning compact latent representations that enable efficient diffusion-based generation across diverse data modalities. The taxonomy reveals four main branches that collectively map the landscape. Autoencoder Architecture and Latent Space Design focuses on the structural choices underlying representation autoencoders—ranging from hierarchical designs like Hierarchical Diffusion Autoencoders[20] to specialized compression schemes such as Deep Compression Autoencoder[13] and adaptive encoding strategies like Adaptive Latent Encoding[29]. Representation Learning and Semantic Encoding emphasizes how autoencoders capture meaningful structure, with works like Diffusion Representation Learner[3] and Diffusion Representation Learning[5] exploring semantic disentanglement and interpretability. Generation Applications Across Domains showcases the breadth of modalities tackled—from language (Latent Diffusion Language[4]) and proteins (Latent Diffusion Protein[6], ProteinAE Diffusion[37]) to DNA sequences (Latent Diffusion DNA[10]) and 3D geometry (Geometric Latent Diffusion[19]). Specialized Applications and Analysis addresses niche use cases and analytical perspectives, including medical imaging (Latent Diffusion Medical[12]) and anomaly detection (Lafite Anomaly Detection[24]).

A particularly active line of work explores the interplay between autoencoder design and diffusion quality, with studies like Improving Diffusability Autoencoders[8] and Diffusion Bridge AutoEncoders[15] investigating how latent space properties affect generative performance. Another contrasting theme is the tension between compression efficiency and semantic fidelity: while Deep Compression Autoencoder[13] and LiteVAE[30] prioritize compact representations, works like Structured Latent Space[18] and Lost Latent Space[33] examine the trade-offs in preserving interpretable structure. The original paper, Diffusion Transformers Autoencoders[0], sits within the Autoencoder Architecture branch alongside MeanFlow Transformers Autoencoders[14], emphasizing transformer-based architectures for representation learning. Compared to neighboring efforts like Diffusion Masked Autoencoders[2], which integrate masking strategies, Diffusion Transformers Autoencoders[0] appears to focus more directly on leveraging transformer expressiveness to refine latent encodings for diffusion, positioning it at the intersection of architectural innovation and representation quality.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. MeanFlow Transformers with Representation Autoencoders

Authors: Zheyuan Hu, Chieh-Hsin Lai, Ge Wu, Yuki Mitsufuji, Stefano Ermon | **Year/Venue**: 2025 | **URL**: View paper

Abstract

MeanFlow (MF) is a diffusion-motivated generative model that enables efficient few-step generation by learning long jumps directly from noise to data. In practice, it is often used as a latent MF by leveraging the pre-trained Stable Diffusion variational autoencoder (SD-VAE) for high-dimensional data modeling. However, MF training remains computationally demanding and is often unstable. During inference, the SD-VAE decoder dominates the generation cost, and MF depends on complex guidance hyperpa...

Relationship Analysis

Both papers belong to the Representation Autoencoders category, using pretrained vision encoders (DINOv2, DINO) with trained decoders to create semantically rich latent spaces for generative modeling. They overlap in addressing the challenge of training diffusion models in high-dimensional RAE latent spaces and both identify issues with naive training approaches. The key difference is that the original paper focuses on adapting Diffusion Transformers (DiTs) to work with RAEs through architectural modifications (DDT head), dimension-dependent noise scheduling, and noise-augmented decoding, while the candidate paper specifically addresses MeanFlow training instability in RAE latent spaces using Consistency Mid-Training, distillation from flow matching teachers, and a two-stage bootstrapping scheme for few-step generation.

Contributions Analysis

Overall novelty summary. The paper proposes Representation Autoencoders (RAEs), which replace traditional VAE encoders with pretrained representation encoders (DINO, SigLIP, MAE) paired with trained decoders. This work resides in the 'Representation Autoencoders' leaf of the taxonomy, which contains only two papers including the original. This sparse population suggests the specific approach of combining pretrained encoders with diffusion transformers is relatively unexplored. The taxonomy shows the broader 'Autoencoder Architecture and Latent Space Design' branch contains five leaves addressing compression, latent properties, masking, unified architectures, and representation autoencoders, indicating moderate activity in autoencoder design overall.

The taxonomy reveals neighboring research directions that contextualize this work. The 'High-Compression and Efficient Autoencoders' leaf (three papers) pursues spatial compression through architectural innovations, while 'Latent Space Properties and Optimization' (four papers) analyzes latent characteristics like smoothness and discriminability. The 'Masked Autoencoder Integration' leaf (one paper) explores masking strategies, and 'Unified End-to-End Architectures' (two papers) merges encoder-decoder-diffusion components. The original paper diverges by emphasizing semantically rich pretrained representations rather than compression ratios or end-to-end unification, carving a distinct niche within the autoencoder design landscape.

Among 21 candidates examined across three contributions, none were found to clearly refute the proposed methods. Contribution A (RAEs) examined 10 candidates with zero refutable matches, suggesting limited direct prior work on this specific encoder-decoder combination. Contribution B (theoretically motivated solutions for high-dimensional diffusion) also examined 10 candidates with no refutations, indicating the theoretical analysis may address gaps in existing literature. Contribution C (DiTDH architecture) examined only one candidate with no overlap. The modest search scope (21 papers) and absence of refutations suggest these contributions occupy relatively novel territory within the examined literature.

Based on the limited search scope of 21 semantically related papers, the work appears to introduce a distinct approach within a sparsely populated research direction. The taxonomy structure confirms that representation autoencoders constitute a small but emerging area, with the original paper and one sibling defining this leaf. However, the analysis does not cover exhaustive literature review or broader architectural surveys, leaving open the possibility of related work outside the top-K semantic matches examined.

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

Contribution 1: Representation Autoencoders (RAEs)

Description: The authors propose Representation Autoencoders (RAEs), which replace traditional VAE encoders with frozen pretrained representation encoders (such as DINOv2, SigLIP, or MAE) paired with lightweight learned decoders. RAEs provide high-quality reconstructions and semantically rich latent spaces while using a scalable transformer-based architecture.

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. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation

URL: View paper

Brief Assessment

RNN Encoder-Decoder[53] focuses on sequence-to-sequence learning for machine translation, not on image reconstruction using pretrained visual encoders with learned decoders as in RAEs.

2. EdgeRunner: Auto-regressive Auto-encoder for Artistic Mesh Generation

URL: View paper

Brief Assessment

EdgeRunner Mesh Generation[55] focuses on auto-regressive mesh generation with a mesh-specific tokenization algorithm, not on replacing VAE encoders with frozen pretrained representation encoders for image reconstruction and diffusion modeling.

3. Inverge: Intelligent visual encoder for bridging modalities in report generation

URL: View paper

Brief Assessment

Inverge Report Generation[57] focuses on medical report generation using a cross-modal query fusion layer between frozen encoders and decoders, not on general-purpose image reconstruction autoencoders for diffusion models.

4. Omni-id: Holistic identity representation designed for generative tasks

URL: View paper

Brief Assessment

Omni-ID Holistic Identity[51] focuses on facial identity representation for generative tasks using a few-to-many reconstruction paradigm with multi-decoder training. The original paper proposes RAEs for general image generation by pairing frozen pretrained encoders (DINOv2, SigLIP, MAE) with learned decoders for diffusion models. These are fundamentally different architectural approaches and application domains.

5. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction

URL: View paper

Brief Assessment

Mofa Face Autoencoder[59] focuses on face reconstruction using a model-based decoder with semantic face parameters (pose, shape, expression, reflectance, illumination), not general-purpose pretrained representation encoders like DINOv2 or SigLIP combined with learned decoders for image reconstruction.

6. Context autoencoder for self-supervised representation learning

URL: View paper

Brief Assessment

Context Autoencoder[52] focuses on masked image modeling for self-supervised learning, not on replacing VAE encoders with frozen pretrained representation encoders for diffusion models. The architectural goals and application domains differ fundamentally.

7. Deep learning for tomographic image reconstruction

URL: View paper

Brief Assessment

Deep Learning Tomographic[54] focuses on tomographic image reconstruction using encoder-decoder networks for medical imaging. The candidate does not address pretrained representation encoders (DINOv2, SigLIP, MAE) combined with learned decoders for generative modeling, which is the core novelty of RAEs.

8. Unsupervised representation learning from pre-trained diffusion probabilistic models

URL: View paper

Brief Assessment

Unsupervised Pretrained Diffusion[56] focuses on adapting pre-trained diffusion models as decoders with learned encoders for autoencoding, not on using frozen pretrained representation encoders (like DINOv2, SigLIP, MAE) as the encoder component of an autoencoder architecture.

9. Unsupervised knowledge-transfer for learned image reconstruction

URL: View paper

Brief Assessment

Unsupervised Knowledge-Transfer[60] addresses unsupervised transfer learning for medical image reconstruction using Bayesian networks, not pretrained representation encoders (DINOv2, SigLIP, MAE) combined with learned decoders for general image reconstruction and generation tasks.

10. End-to-End Object Detection with Transformers

URL: View paper

Brief Assessment

DETR Object Detection[58] focuses on object detection using transformers with set prediction, not on image reconstruction or autoencoder architectures for generative modeling.

Contribution 2: Theoretically motivated solutions for high-dimensional diffusion

Description: The authors identify and address three key challenges in enabling diffusion transformers to operate effectively in high-dimensional RAE latent spaces: transformer width must match token dimensionality, noise scheduling must be dimension-dependent, and decoders require noise-augmented training. These solutions are supported by theoretical analysis and empirical validation.

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. Continuous-time Discrete-space Diffusion Model for Recommendation

URL: View paper

Brief Assessment

Continuous-time Discrete-space Diffusion[69] operates in discrete state-space for recommendation tasks, not high-dimensional continuous latent spaces. The candidate focuses on masking operations and discrete item transitions, while the original addresses continuous-space challenges like transformer width scaling and dimension-dependent noise scheduling for image generation.

2. Renormalization group flow, optimal transport, and diffusion-based generative model.

URL: View paper

Brief Assessment

Renormalization Group Flow[67] focuses on applying RG flow in Fourier space for diffusion models, not on addressing high-dimensional latent space challenges in diffusion transformers. The candidate does not discuss transformer width requirements, dimension-dependent noise scheduling for high-dimensional tokens, or noise-augmented decoder training.

3. Ant: Adaptive noise schedule for time series diffusion models

URL: View paper

Brief Assessment

Ant Adaptive Noise[65] focuses on adaptive noise scheduling for time series diffusion models based on non-stationarity statistics, not on high-dimensional latent space challenges or dimension-dependent architectural requirements for diffusion transformers.

4. Priority-Centric Human Motion Generation in Discrete Latent Space

URL: View paper

Brief Assessment

Priority-Centric Motion Generation[66] focuses on text-to-motion generation in discrete latent spaces with priority-based noise scheduling for motion tokens, not on the general challenges of high-dimensional diffusion in RAE latent spaces or dimension-dependent noise scheduling for image generation.

5. Trans-Dimensional Generative Modeling via Jump Diffusion Models

URL: View paper

Brief Assessment

Jump Diffusion Models[63] addresses trans-dimensional generation (varying number of components) rather than high-dimensional fixed-token diffusion. The technical focus differs fundamentally from dimension-dependent noise scheduling and transformer width requirements.

6. Rethinking the noise schedule of diffusion-based generative models

URL: View paper

Brief Assessment

Rethinking Noise Schedule[68] focuses on noise scheduling strategies through power spectrum analysis and proposes WSNR metrics for different resolutions. It does not address the specific challenges of high-dimensional RAE latent spaces, transformer width requirements matching token dimensionality, or noise-augmented decoder training that the original paper tackles.

7. Schedule On the Fly: Diffusion Time Prediction for Faster and Better Image Generation

URL: View paper

Brief Assessment

Schedule On Fly[64] focuses on adaptive noise scheduling across prompts for text-to-image generation, not on addressing challenges of high-dimensional latent spaces or dimension-dependent noise scheduling for diffusion transformers.

8. Diffusion-based Large Language Models Survey

URL: View paper

Brief Assessment

Diffusion LLM Survey[62] discusses diffusion models in language contexts with token embeddings and noise schedules, but does not address the specific challenges of high-dimensional RAE latent spaces (transformer width matching token dimensionality, dimension-dependent noise scheduling, noise-augmented decoder training) that the original paper identifies and solves.

9. Diffusion models with learned adaptive noise

URL: View paper

Brief Assessment

Adaptive Noise Diffusion[61] focuses on learning multivariate noise schedules to improve log-likelihood estimation in diffusion models, not on enabling diffusion transformers to operate in high-dimensional RAE latent spaces with dimension-dependent noise scheduling and decoder training strategies.

10. Channel-wise Noise Scheduled Diffusion for Inverse Rendering in Indoor Scenes

URL: View paper

Brief Assessment

Channel-wise Noise Scheduled[70] focuses on inverse rendering with channel-wise noise scheduling for decomposing RGB images into geometry, material, and lighting. This is fundamentally different from the original paper's work on enabling diffusion transformers in high-dimensional RAE latent spaces with dimension-dependent noise scheduling for image generation.

Contribution 3: DiTDH architecture with wide DDT head

Description: The authors introduce DiTDH, an augmented DiT architecture that incorporates a wide, shallow transformer module (DDT head) dedicated to denoising. This design provides sufficient model width for high-dimensional diffusion without scaling the entire backbone, achieving faster convergence and state-of-the-art generation performance.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. TinyFusion: Diffusion Transformers Learned Shallow

URL: View paper

Brief Assessment

TinyFusion Shallow[71] focuses on depth pruning to remove redundant layers from diffusion transformers, not on adding wide shallow heads for denoising. The architectural approaches are fundamentally different.

Appendix: Text Similarity Detection

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

References

- [0] Diffusion Transformers with Representation Autoencoders View paper
- [1] Latent diffusion modeling of porous media informed by spatial statistics. View paper
- [2] Diffusion models as masked autoencoders View paper
- [3] Diffusion model as representation learner View paper
- [4] Latent diffusion for language generation View paper
- [5] Diffusion based representation learning View paper
- [6] A latent diffusion model for protein structure generation View paper
- [7] Factorized diffusion autoencoder for unsupervised disentangled representation learning View paper
- [8] Improving the diffusability of autoencoders View paper
- [9] Lion: Latent point diffusion models for 3d shape generation View paper
- [10] Latent diffusion model for dna sequence generation View paper
- [11] Latent Diffusion Models with Masked AutoEncoders View paper
- [12] Latent Diffusion Autoencoders: Toward Efficient and Meaningful Unsupervised Representation Learning in Medical Imaging View paper
- [13] Deep compression autoencoder for efficient high-resolution diffusion models View paper
- [14] MeanFlow Transformers with Representation Autoencoders View paper
- [15] Diffusion Bridge AutoEncoders for Unsupervised Representation Learning View paper
- [16] Aeroblade: Training-free detection of latent diffusion images using autoencoder reconstruction error View paper
- [17] Latent diffusion for neural spiking data View paper
- [18] Dc-ae 1.5: Accelerating diffusion model convergence with structured latent space View paper

- [19] Geometric Latent Diffusion Models for 3D Molecule Generation View paper
- [20] Hierarchical diffusion autoencoders and disentangled image manipulation View paper
- [21] Diffusion As Self-Distillation: End-to-End Latent Diffusion In One Model View paper
- [22] Crystal diffusion variational autoencoder for periodic material generation View paper
- [23] Dae-talker: High fidelity speech-driven talking face generation with diffusion autoencoder View paper
- [24] Lafite: Latent diffusion model with feature editing for unsupervised multi-class anomaly detection View paper
- [25] H-space sparse autoencoders View paper
- [26] Latent 3d graph diffusion View paper
- [27] Diffusion autoencoders: Toward a meaningful and decodable representation View paper
- [28] High-Resolution Image Synthesis with Latent Diffusion Models View paper
- [29] Unified generation, reconstruction, and representation: generalized diffusion with adaptive latent encoding-decoding View paper
- [30] Litevae: Lightweight and efficient variational autoencoders for latent diffusion models View paper
- [31] Executing your commands via motion diffusion in latent space View paper
- [32] Causal Diffusion Autoencoders: Toward Counterfactual Generation via Diffusion Probabilistic Models View paper
- [33] Lost in Latent Space: An Empirical Study of Latent Diffusion Models for Physics Emulation View paper
- [34] cDVAE: Multimodal generative conditional diffusion guided by variational autoencoder latent embedding for virtual 6D phase space diagnostics View paper
- [35] Mixed-Type Tabular Data Synthesis with Score-based Diffusion in Latent Space View paper
- [36] DiffSDA: Unsupervised Diffusion Sequential Disentanglement Across Modalities View paper
- [37] ProteinAE: Protein Diffusion Autoencoders for Structure Encoding View paper
- [38] TimeAutoDiff: Combining Autoencoder and Diffusion model for time series tabular data synthesizing View paper
- [39] Pluralistic Aging Diffusion Autoencoder View paper
- [40] Representation learning with diffusion models View paper
- [41] GLDM: hit molecule generation with constrained graph latent diffusion model View paper
- [42] LN3DIFF++: Scalable Latent Neural Fields Diffusion for Speedy 3D Generation. View paper
- [43] Autodecoding latent 3d diffusion models View paper
- [44] Timeldm: Latent diffusion model for unconditional time series generation View paper
- [45] DGAE: Diffusion-Guided Autoencoder for Efficient Latent Representation Learning View paper
- [46] Automated Learning of Semantic Embedding Representations for Diffusion Models View paper
- [47] Masked autoencoders are effective tokenizers for diffusion models View paper
- [48] RoSteALS: Robust Steganography using Autoencoder Latent Space View paper
- [49] Binary latent diffusion View paper
- [50] MM-LDM: Multi-Modal Latent Diffusion Model for Sounding Video Generation View paper
- [51] Omni-id: Holistic identity representation designed for generative tasks View paper
- [52] Context autoencoder for self-supervised representation learning View paper
- [53] Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation View paper
- [54] Deep learning for tomographic image reconstruction View paper
- [55] EdgeRunner: Auto-regressive Auto-encoder for Artistic Mesh Generation View paper
- [56] Unsupervised representation learning from pre-trained diffusion probabilistic models View paper
- [57] Inverge: Intelligent visual encoder for bridging modalities in report generation View paper
- [58] End-to-End Object Detection with Transformers View paper
- [59] Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction View paper
- [60] Unsupervised knowledge-transfer for learned image reconstruction View paper
- [61] Diffusion models with learned adaptive noise View paper
- [62] Diffusion-based Large Language Models Survey View paper
- [63] Trans-Dimensional Generative Modeling via Jump Diffusion Models View paper
- [64] Schedule On the Fly: Diffusion Time Prediction for Faster and Better Image Generation View paper
- [65] Ant: Adaptive noise schedule for time series diffusion models View paper
- [66] Priority-Centric Human Motion Generation in Discrete Latent Space View paper
- [67] Renormalization group flow, optimal transport, and diffusion-based generative model. View paper
- [68] Rethinking the noise schedule of diffusion-based generative models View paper
- [69] Continuous-time Discrete-space Diffusion Model for Recommendation View paper
- [70] Channel-wise Noise Scheduled Diffusion for Inverse Rendering in Indoor Scenes View paper
- [71] TinyFusion: Diffusion Transformers Learned Shallow View paper