Novelty Assessment Report

Paper: PixNerd: Pixel Neural Field Diffusion
PDF URL: https://openreview.net/pdf?id=BDnOrExHmt
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

The current success of diffusion transformers are built on the compressed latent space shaped by the pre-trained variational autoencoder(VAE). However, this two-stage training paradigm inevitably introduces accumulated errors and decoding artifacts. To avoid these problems, researchers return to pixel space modeling but at the cost of complicated cascade pipelines and increased token complexity. Motivated by the simple yet effective diffusion transformer architectures on the latent space, we propose to model pixel space diffusion using a large-patch diffusion transformer and employ neural fields to decode these large patches, leading to a single-stage streamlined end-to-end solution, which we coin as pixel neural field diffusion transformer (**PixNerd**). Thanks to the efficient neural field representation in PixNerd, we achieve **1.93 FID** on ImageNet 256x256 and nearly **8x lower latency** without any complex cascade pipeline or VAE. We also extend our PixNerd framework to text-to-image applications. Our PixNerd-XXL/16 achieves a competitive 0.73 overall score on the GenEval benchmark and 80.9 overall score on the DPG benchmark.

Core Task Landscape

This paper addresses: **Pixel Space Diffusion Modeling with Neural Fields**

A total of **26 papers** were analyzed and organized into a taxonomy with **14 categories**.

Taxonomy Overview

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

- **Direct Pixel-Space Neural Field Generation**
- **Latent-Space Neural Field Diffusion**
- **Neural Radiance Field Editing and Synthesis**
- **Texture Synthesis on 3D Surfaces with Diffusion**
- **Diffusion Models on Continuous Function Spaces**
- **Domain-Specific Neural Field Diffusion Applications**
- **Neural Field Dynamics and Theoretical Foundations**

Complete Taxonomy Tree

- Pixel Space Diffusion Modeling with Neural Fields Survey Taxonomy
- Direct Pixel-Space Neural Field Generation
  - End-to-End Pixel Diffusion with Neural Field Decoding ★ (2 papers)
  - [0] PixNerd: Pixel Neural Field Diffusion (Anon et al., 2026) View paper
  - [18] Image Neural Field Diffusion Models (Yinbo Chen, 2024) View paper
- Latent-Space Neural Field Diffusion
  - 3D Shape and Geometry Latent Diffusion (4 papers)
  - [1] 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models (Biao Zhang, 2023) View paper
  - [6] LN3Diff: Scalable Latent Neural Fields Diffusion for Speedy 3D Generation (Lan, 2024) View paper
  - [13] Part-aware Shape Generation with Latent 3D Diffusion of Neural Voxel Fields (Yuhang Huang, 2025) View paper
  - [23] 3D Neural Field Generation Using Triplane Diffusion (J. Ryan Shue, 2023) View paper
  - Scene-Level Latent Diffusion with Neural Fields (1 papers)
  - [11] Neuralfield-ldm: Scene generation with hierarchical latent diffusion models (Seung Wook Kim, 2023) View paper
  - Domain-Agnostic Implicit Neural Representation Diffusion (2 papers)
  - [16] DDMI: Domain-Agnostic Latent Diffusion Models for Synthesizing High-Quality Implicit Neural Representations (Kim, 2024) View paper
  - [24] HyperDiffusion: Generating Implicit Neural Fields with Weight-Space Diffusion (Ziya ErkoÃ§, 2023) View paper
- Neural Radiance Field Editing and Synthesis
  - Text-Guided NeRF Editing with Diffusion Priors (2 papers)
  - [2] ViCA-NeRF: View-Consistency-Aware 3D Editing of Neural Radiance Fields (Dong, 2024) View paper
  - [3] Dreameditor: Text-driven 3d scene editing with neural fields (Jingyu Zhuang, 2023) View paper
  - NeRF View Extrapolation and Inpainting with Diffusion (2 papers)
  - [9] Taming latent diffusion model for neural radiance field inpainting (Lin, 2024) View paper
  - [12] ExtraNeRF: Visibility-Aware View Extrapolation of Neural Radiance Fields with Diffusion Models (Meng-Li Shih, 2024) View paper
  - NeRF Super-Resolution with Diffusion Guidance (1 papers)
  - [14] Advancing Super-Resolution in Neural Radiance Fields via Variational Diffusion Strategies (Shrey Vishen, 2024) View paper
  - NeRF Regularization with Diffusion Priors (2 papers)
  - [8] ID-NeRF: Indirect diffusion-guided neural radiance fields for generalizable view synthesis (YaoKun Li, 2024) View paper

    - [22] DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models (Jamie Wynn, 2023) View paper
- Texture Synthesis on 3D Surfaces with Diffusion (2 papers)
    - [4] Texfusion: Synthesizing 3d textures with text-guided image diffusion models (Tianshi Cao, 2023) View paper
    - [7] Single mesh diffusion models with field latents for texture generation (Thomas W. Mitchel, 2024) View paper
- Diffusion Models on Continuous Function Spaces (1 papers)
    - [5] Diffusion probabilistic fields (Zhuang, 2023) View paper
- Domain-Specific Neural Field Diffusion Applications
    - Spatiotemporal Turbulence Simulation with Neural Field Diffusion (3 papers)
    - [17] Conditional Neural Field Latent Diffusion Model Generating Spatiotemporal Turbulence (Du Pan, 2024) View paper
    - [19] CoNFiLD: Conditional Neural Field Latent Diffusion Model Generating Spatiotemporal Turbulence (Du Pan, 2024) View paper
    - [21] Conditional neural field latent diffusion model for generating spatiotemporal turbulence (Pan Du, 2024) View paper
    - Medical Imaging Reconstruction with Neural Field Diffusion (2 papers)
    - [15] Diff-inr: Generative regularization for electrical impedance tomography (Tong, 2024) View paper
    - [20] Highly accelerated MRI via implicit neural representation guided posterior sampling of diffusion models. (Jiayue Chu, 2024) View paper
    - Light Field Parameterization with Diffusion Equations (1 papers)
    - [10] Diffusion equation based parameterization of light field and computational imaging model (Chang Liu, 2022) View paper
- Neural Field Dynamics and Theoretical Foundations (2 papers)
    - [25] Power spectrum and diffusion of the Amari neural field (Luca Salasnich, 2022) View paper
    - [26] Neural Field Models with Transmission Delays and Diffusion (Len Spek, 2022) View paper

Narrative

Core task: pixel space diffusion modeling with neural fields. This emerging area combines diffusion generative models with neural field representations to synthesize and manipulate continuous visual data. The taxonomy reveals several complementary directions. Direct Pixel-Space Neural Field Generation methods, such as PixNerd[0] and Image Neural Field[18], learn to generate neural field parameters end-to-end while diffusing in pixel or coordinate space, enabling flexible resolution and continuous outputs. Latent-Space Neural Field Diffusion approaches like NeuralField-LDM[11] and HyperDiffusion[24] instead encode neural fields into compact latent codes before applying diffusion, trading some direct pixel control for efficiency and scalability. Neural Radiance Field Editing and Synthesis (e.g., DreamEditor[3], ViCA-NeRF[2]) focuses on manipulating 3D scene representations, while Texture Synthesis on 3D Surfaces with Diffusion (TexFusion[4], Single Mesh Diffusion[7]) targets surface appearance. Diffusion Models on Continuous Function Spaces (Diffusion Probabilistic Fields[5], Diff-INR[15]) explore the theoretical grounding of diffusing over function-valued distributions, and Domain-Specific Neural Field Diffusion Applications extend these ideas to medical imaging (Accelerated MRI[20]) and scientific data (Spatiotemporal Turbulence[21]).

A central tension runs between end-to-end pixel-space methods and latent-space strategies: the former preserve fine-grained control and interpretability, while the latter achieve faster sampling and better scalability for high-dimensional scenes. PixNerd[0] exemplifies the direct pixel-space philosophy, generating neural field weights through a diffusion process that operates close to the final rendering, much like Image Neural Field[18]. This contrasts with latent approaches such as NeuralField-LDM[11], which compress neural fields into lower-dimensional codes before diffusion, sacrificing some pixel-level transparency for computational gains. Meanwhile, works like Diffusion Probabilistic Fields[5] provide a rigorous function-space perspective that underpins both paradigms. PixNerd[0] sits squarely in the Direct Pixel-Space branch, emphasizing end-to-end pixel diffusion with neural field decoding, and shares conceptual ground with Image Neural Field[18] in prioritizing continuous, resolution-agnostic generation without an intermediate latent bottleneck.

Related Works in Same Category

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

1. Image Neural Field Diffusion Models

Authors: Yinbo Chen, Oliver Wang, Richard Zhang, Eli Shechtman, Xiaolong Wang, et al. (6 authors total) | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Diffusion models have shown an impressive ability to model complex data distributions, with several key advantages over GANs, such as stable training, better coverage of the training distribution's modes, and the ability to solve inverse problems without extra training. However, most diffusion models learn the distribution of fixed-resolution images. We propose to learn the distribution of continuous images by training diffusion models on image neural fields, which can be rendered at any resolut...

Relationship Analysis

Both papers belong to the End-to-End Pixel Diffusion with Neural Field Decoding category, employing single-stage frameworks that model diffusion in pixel space and decode via neural fields. They overlap in using neural field representations to handle pixel-space diffusion without cascades or VAEs, both aiming to generate continuous/high-quality images directly. The key difference is that PixNerd uses patch-wise adaptive neural field heads with predicted MLP weights from diffusion transformer features for large-patch decoding, while the candidate paper focuses on learning distributions of continuous image neural fields that can be rendered at arbitrary resolutions, converting latent diffusion autoencoders into image neural field autoencoders and training on mixed-resolution datasets.

Contributions Analysis

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

Contribution 1: PixNerd: Pixel Neural Field Diffusion Transformer

Description: The authors introduce PixNerd, a novel architecture that combines large-patch diffusion transformers with neural field decoding for pixel-space image generation. This approach replaces the traditional linear projection with a patch-wise implicit neural field head, enabling efficient single-stage end-to-end training without requiring VAEs or cascade pipelines.

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. Latent Diffusion Transformer with Local Neural Field as PDE Surrogate Model

URL: View paper

Brief Assessment

Latent Diffusion Transformer[38] focuses on PDE surrogate modeling for scientific simulations (fluid dynamics, Burgers' equation), not pixel-space image generation. The architectures serve fundamentally different purposes despite both using neural fields with transformers.

Contribution 2: Patch-wise adaptive neural field head for large-patch decoding

Description: The authors design a patch-wise adaptive neural field head whose weights are predicted by the diffusion transformer's last hidden features. For each pixel within a patch, local coordinates and noisy pixel values are encoded and fed into a neural field MLP to predict diffusion velocity, addressing the challenge of learning fine details with large-patch configurations.

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. NITO: Neural Implicit Fields for Resolution-free and Domain-Adaptable Topology Optimization

URL: View paper

Brief Assessment

NITO[37] focuses on topology optimization using neural implicit fields for structural design, not diffusion models for image generation. The neural field usage is fundamentally different from the original paper's patch-wise adaptive head for diffusion transformers.

2. CoNFiLD: Conditional Neural Field Latent Diffusion Model Generating Spatiotemporal Turbulence

URL: View paper

Brief Assessment

CoNFiLD[19] uses conditional neural fields for encoding spatiotemporal turbulence data into latent space, not for patch-wise decoding in diffusion transformers. The architectural purposes and application domains differ fundamentally.

Contribution 3: Competitive performance on class-to-image and text-to-image benchmarks

Description: The authors demonstrate that PixNerd achieves competitive results on both class-conditional and text-to-image generation tasks. On ImageNet 256×256, PixNerd-XL/16 obtains 1.93 FID with computational demands similar to latent diffusion models, while PixNerd-XXL/16 achieves strong scores on GenEval and DPG benchmarks for text-to-image generation.

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. Text-to-image diffusion models are zero shot classifiers

URL: View paper

Brief Assessment

Zero Shot Classifiers[30] focuses on using diffusion models as zero-shot classifiers for discriminative tasks, not on pixel-space diffusion model performance for generative tasks. The candidate evaluates classification accuracy on datasets like ImageNet, while the original paper reports FID scores for image generation quality.

2. Diffusion models without attention

URL: View paper

Brief Assessment

Diffusion Without Attention[31] focuses on class-conditional ImageNet generation using state space models instead of attention mechanisms, not on pixel-space diffusion with neural fields for both class-to-image and text-to-image tasks as in PixNerd.

3. Palette: Image-to-image diffusion models

URL: View paper

Brief Assessment

Palette[29] focuses on image-to-image translation tasks (colorization, inpainting, uncropping, JPEG restoration) rather than class-conditional or text-to-image generation benchmarks like ImageNet or GenEval/DPG that PixNerd targets.

4. Aid: Attention interpolation of text-to-image diffusion

URL: View paper

Brief Assessment

Aid[35] focuses on text-to-image interpolation and conditional generation tasks, not on pixel-space diffusion model performance benchmarks like ImageNet or GenEval that PixNerd addresses.

5. Spatext: Spatio-textual representation for controllable image generation

URL: View paper

Brief Assessment

Spatext[33] focuses on controllable image generation with spatio-textual representations for scene control, not on pixel-space diffusion model performance benchmarks like ImageNet or text-to-image generation metrics (FID, GenEval, DPG).

6. Diffit: Diffusion vision transformers for image generation

URL: View paper

Brief Assessment

Diffit[34] focuses on architectural innovations (time-dependent self-attention in transformers) rather than pixel-space diffusion. It achieves 1.73 FID on ImageNet-256 in latent space, not pixel space like PixNerd.

7. PixArt-: Fast Training of Diffusion Transformer for Photorealistic Text-to-Image Synthesis

URL: View paper

Brief Assessment

PixArt[32] focuses on text-to-image generation using diffusion transformers with VAE compression, achieving FID scores on COCO and T2I-CompBench. The original paper addresses pixel-space diffusion without VAE, targeting ImageNet class-conditional generation and different text-to-image benchmarks (GenEval, DPG). These are fundamentally different approaches and evaluation contexts.

8. DeCo: Frequency-Decoupled Pixel Diffusion for End-to-End Image Generation

URL: View paper

Prior Art Analysis

DeCo[36] demonstrates that pixel-space diffusion models can achieve competitive or superior performance on ImageNet and text-to-image benchmarks before PixNerd's publication. DeCo[36] achieves 1.62 FID on ImageNet 256×256 and 2.22 FID on ImageNet 512×512, which are better than PixNerd's reported 1.93 FID. On text-to-image generation, DeCo[36] achieves 0.86 overall score on

GenEval and 81.4 on DPG-bench, surpassing PixNerd-XXL/16's 0.73 GenEval score and 80.9 DPG score. This demonstrates that competitive pixel-space diffusion performance on these benchmarks was already achieved prior to PixNerd's work.

Evidence

Evidence 1 - **Rationale**: DeCo[36] achieves a higher GenEval score (0.86 vs 0.73), demonstrating superior text-to-image performance on the same benchmark before PixNerd's publication. - **Original**: our pixnerd-xxl/16 achieves a competitive 0.73 overall score on the geneval benchmark and 80.9 overall score on the dpg benchmark. - **Candidate**: our pretrained text-to-image model achieves a leading overall score of 0.86 on geneval in system-level comparison.

Evidence 2 - **Rationale**: DeCo[36] explicitly states achieving superior performance among pixel diffusion models with better FID scores than PixNerd on ImageNet benchmarks. - **Original**: for the class-to-image generation, on imagenet256x256, our pixnerd-xl/16 achieves a competitive 1.93 fid with similar computation demands as its latent counterpart. - **Candidate**: deco achieves superior performance among pixel diffusion models, attaining fid of 1.62 (256x256) and 2.22 (512x512) on imagenet, closing the gap with latent diffusion methods.

Evidence 3 - **Rationale**: DeCo[36] achieves higher scores on both GenEval (0.86 vs 0.73) and DPG-bench (81.4 vs 80.9), demonstrating superior text-to-image performance on the same benchmarks. - **Original**: for the text-to-image generation, our pixnerd-xxl/16 achieves a 0.73 overall score on the geneval benchmark and 80.9 average score on dpg benchmark. - **Candidate**: our pretrained text-to-image model also achieves leading results on geneval (0.86) and dpg-bench (81.4) in system-level evaluation.

---

9. On distillation of guided diffusion models

URL: View paper

Brief Assessment

Distillation Guided Diffusion[27] focuses on distilling classifier-free guided diffusion models to reduce sampling steps, not on developing a new pixel-space diffusion architecture. The paper demonstrates performance on ImageNet and text-to-image tasks but through a distillation approach rather than the neural field-based pixel diffusion framework proposed in the original paper.

---

10. All are worth words: A vit backbone for diffusion models

URL: View paper

Brief Assessment

ViT Backbone[28] focuses on Vision Transformer architectures for diffusion models with U-Net-like skip connections, achieving strong results on ImageNet and MS-COCO. However, it does not challenge PixNerd's novelty of using neural fields for large-patch pixel-space diffusion without VAE.

Appendix: Text Similarity Detection

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

References

- [0] PixNerd: Pixel Neural Field Diffusion View paper
- [1] 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models View paper
- [2] ViCA-NeRF: View-Consistency-Aware 3D Editing of Neural Radiance Fields View paper
- [3] Dreameditor: Text-driven 3d scene editing with neural fields View paper
- [4] Texfusion: Synthesizing 3d textures with text-guided image diffusion models View paper
- [5] Diffusion probabilistic fields View paper
- [6] LN3Diff: Scalable Latent Neural Fields Diffusion for Speedy 3D Generation View paper
- [7] Single mesh diffusion models with field latents for texture generation View paper
- [8] ID-NeRF: Indirect diffusion-guided neural radiance fields for generalizable view synthesis View paper
- [9] Taming latent diffusion model for neural radiance field inpainting View paper
- [10] Diffusion equation based parameterization of light field and computational imaging model View paper
- [11] Neuralfield-ldm: Scene generation with hierarchical latent diffusion models View paper
- [12] ExtraNeRF: Visibility-Aware View Extrapolation of Neural Radiance Fields with Diffusion Models View paper
- [13] Part-aware Shape Generation with Latent 3D Diffusion of Neural Voxel Fields View paper
- [14] Advancing Super-Resolution in Neural Radiance Fields via Variational Diffusion Strategies View paper
- [15] Diff-inr: Generative regularization for electrical impedance tomography View paper
- [16] DDMI: Domain-Agnostic Latent Diffusion Models for Synthesizing High-Quality Implicit Neural Representations View paper
- [17] Conditional Neural Field Latent Diffusion Model Generating Spatiotemporal Turbulence View paper
- [18] Image Neural Field Diffusion Models View paper
- [19] CoNFiLD: Conditional Neural Field Latent Diffusion Model Generating Spatiotemporal Turbulence View paper
- [20] Highly accelerated MRI via implicit neural representation guided posterior sampling of diffusion models. View paper
- [21] Conditional neural field latent diffusion model for generating spatiotemporal turbulence View paper
- [22] DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models View paper
- [23] 3D Neural Field Generation Using Triplane Diffusion View paper
- [24] HyperDiffusion: Generating Implicit Neural Fields with Weight-Space Diffusion View paper
- [25] Power spectrum and diffusion of the Amari neural field View paper
- [26] Neural Field Models with Transmission Delays and Diffusion View paper
- [27] On distillation of guided diffusion models View paper
- [28] All are worth words: A vit backbone for diffusion models View paper
- [29] Palette: Image-to-image diffusion models View paper
- [30] Text-to-image diffusion models are zero shot classifiers View paper
- [31] Diffusion models without attention View paper
- [32] PixArt-: Fast Training of Diffusion Transformer for Photorealistic Text-to-Image Synthesis View paper
- [33] Spatext: Spatio-textual representation for controllable image generation View paper
- [34] Diffit: Diffusion vision transformers for image generation View paper
- [35] Aid: Attention interpolation of text-to-image diffusion View paper
- [36] DeCo: Frequency-Decoupled Pixel Diffusion for End-to-End Image Generation View paper
- [37] NITO: Neural Implicit Fields for Resolution-free and Domain-Adaptable Topology Optimization View paper
- [38] Latent Diffusion Transformer with Local Neural Field as PDE Surrogate Model View paper