← Back to **Author Console** (/group?id=thecvf.com/ICCV/2025/Workshop/SP4V/Authors#your-submissions)

GaussianDiffusion: Learning Image Generation Process in GaussianRepresentation Space

PDF (/pdf?id=btllSD53lf)

Simon Coessens (/profile?id=~Simon_Coessens1), Arijit Samal (/profile?id=~Arijit_Samal1), Akash Malhotra (/profile?id=~Akash_Malhotra1), Nacera seghouani (/profile?id=~Nacera_seghouani1)

18 Jul 2025 (modified: 20 Aug 2025) Submitted to SP4V SP4V, Reviewers, Authors Revisions (/revisions?id=btllSD53lf) BibTeX

Keywords: Diffusion Models, Structured Latent Representations, 2D Gaussian Splatting, Image generation, Transformer, Gaussian Image Representation, Low-compute

TL;DR: We introduce GaussianDiffusion, a diffusion model operating on 2D Gaussian splats, offering faster, interpretable image generation with fewer steps than pixel-based baselines.

Abstract:
Diffusion models have become a leading approach in generative image modeling, but many still operate in dense pixel space, a representation that is computationally intensive and lacks geometric structure. We propose GaussianDiffusion, a framework that performs the denoising process entirely in a latent space composed of 2D Gaussians. Each image is encoded as a set of 150 anisotropic Gaussian splats, parameterized by position, covariance, and color. To model their dynamics, we introduce GaussianTransformer, a permutation-equivariant transformer that serves as the denoising network. Evaluated on MNIST and Sprites datasets, our method achieves visual quality comparable to a pixel space U-Net baseline, while reducing the number of sampling steps from 1000 to 200 and the per-step cost from 11.4 GFLOPs to 4 GFLOPs, resulting in an overall 22× improvement in generation time on an A100 GPU. In contrast to latent diffusion models, our approach does not require an auxiliary autoencoder and preserves full editability of the latent. These findings suggest that structured geometric representations can offer efficient and interpretable alternatives to latent and pixel-based diffusion.

Submission Number: 21

Filter by reply type... Filter by author... Search keywords... Sort: Newest First

Everyone | Program Chairs | Submission21 Reviewers | Submission21 Authors | Submission21... | Submission21... | Submission21...

3 / 3 replies shown

Add: **Withdrawal**

Paper Decision

Decision by Program Chairs 19 Aug 2025, 18:16 Program Chairs, Reviewers, Authors

Decision: Reject

Comment:
The reviewers find the idea of using Gaussian splatting in diffusion models novel and potentially efficient. However, they raised concerns about the limited experiments on simple datasets, lack of strong baseline comparisons and ablations, and underdeveloped writing and presentation. While the concept is promising, the paper is not yet ready for publication, and the PC encourages the authors to strengthen experiments and comparisons for a future venue.

Official review

Official Review by Reviewer s4W8 16 Aug 2025, 05:37 (modified: 20 Aug 2025, 01:23) Program Chairs, Reviewers Submitted, Reviewer s4W8, Authors Revisions (/revisions?id=aIFvKgv0SO)

Review:
This paper proposes a novel splatting-based diffusion model. It demonstrates its efficiency in reducing the number of sampling steps without training a VAE model.

> Strengths:

- The proposed method is very interesting and novel. I think this is a promising direction.

> Weaknesses:

- The paper lacks ablation studies. In addition, the method introduces several new hyperparameters, but it is unclear how these hyperparameters are selected. For example, why are 150 anisotropic Gaussian splats used?
- The results still need further improvement. The paper only shows ability on very simple MNIST generation, and the generated Sprites lack detail and the quality is poor. I think the paper needs more experiments to demonstrate the effectiveness of the proposed method.

Rating: 5: Marginally below acceptance threshold
Confidence: 3: The reviewer is fairly confident that the evaluation is correct

Promising concept with efficient design, but limited validation and underdeveloped presentation

Official Review by Reviewer tbmu 28 Jul 2025, 15:40 (modified: 20 Aug 2025, 01:23) Program Chairs, Reviewers Submitted, Reviewer tbmu, Authors Revisions (/revisions?id=Fa5t1egkTQ)

Review:
Summary

This paper proposes Gaussian Diffusion, a novel framework for image generation that operates in a latent space composed of 2D anisotropic Gaussian splats, rather than the conventional dense pixel space. The authors argue that this representation is both more computationally efficient and geometrically meaningful. Each image is encoded as a set of 150 Gaussian splats, parameterized by position, covariance, and color. To model the denoising dynamics, the authors introduce a Gaussian Transformer, a permutation-equivariant transformer that replaces the traditional U-Net architecture. Unlike latent diffusion models, this approach does not require an autoencoder and retains full editability in the latent space. The method is evaluated on MNIST and Sprites datasets, demonstrating significant improvements in sampling speed (22× faster on A100) and computational cost (from 11.4 GFLOPs to 4 GFLOPs per step), while reducing the number of sampling steps from 1000 to 200.

Strengths

1. The integration of anisotropic Gaussian splatting into the diffusion framework is novel and enables efficient rendering in O(n) time, offering a substantial advantage in terms of speed and resource usage.
2. The proposed representation introduces a level of geometric structure that is typically absent in pixel-based diffusion models, opening up new possibilities for interpretable and editable generation.
3. The design of the Gaussian Transformer is well-motivated and technically reasonable, aligning well with the permutation-invariant nature of the Gaussian set representation.

Weaknesses and Questions

1. The datasets used (MNIST and Sprites) are relatively simple and do not convincingly demonstrate the benefits of geometric modeling. It remains unclear whether the proposed method offers tangible advantages over latent diffusion models in more complex settings where geometric priors are meaningful. A more compelling use case might be video generation, where object dynamics and temporal coherence are critical.
2. While the paper emphasizes computational efficiency, it lacks a thorough comparison in terms of generation quality against strong baselines like U-Net-based diffusion or latent diffusion models. Without this, it is difficult to assess the overall effectiveness of the proposed method.
3. Although this is a workshop paper, the structure and writing feel more like a draft. The presentation lacks polish and depth, which affects the clarity and professionalism of the submission.

Overall Assessment

The concept of GaussianDiffusion is fresh and promising, particularly in terms of efficiency. However, the current version of the paper lacks strong experimental support and suffers from presentation issues. The idea is suitable for workshop discussion, but the paper would benefit from stronger experiments. I would recommend a borderline score, slightly leaning toward rejection, unless these issues are addressed.

Rating: 5: Marginally below acceptance threshold
Confidence: 3: The reviewer is fairly confident that the evaluation is correct