
Forward Facing 3D Gaussian Splatting as Markov Chain Monte Carlo

Jeongtaek Oh^{* 1} Wonjun Jeong^{* 2} Huijeong Choe^{* 2} Sanghyun Hahn^{* 3}

Abstract

Novel view synthesis (NVS) has recently gained attention for predicting target images with arbitrary camera pose from novel viewpoints given multiple perspectives to serve various applications such as augmented/virtual reality, and 3D content creation. A recent approach, 3D Gaussian Splatting (3DGS), has illustrated remarkable potential for fast training and rendering by utilizing 3D Gaussian representations. However, they rely heavily on heuristic-based Adaptive Density Control (ADC), which can limit performance, particularly in challenging forward-facing scenarios. In this work, we present Forward Facing 3D Gaussian Splatting as Markov Chain Monte Carlo (FF-3DGS-MCMC), an enhanced NVS method designed specifically for forward-facing scenes. Built upon the 3DGS as Markov Chain Monte Carlo framework, our approach introduces depth map initialization to improve scene coverage and mitigate performance degradation. Furthermore, we refine the heuristic ADC process and propose a novel regularization scheme that helps prevent overfitting, improving rendering quality. Experimental results show that FF-3DGS-MCMC improves forward-facing NVS, offering higher-quality scene reconstructions.

1. Introduction

Novel View Synthesis (NVS) is a crucial task in computer vision, which involves predicting images from unseen viewpoints based on images captured from multiple perspectives. Recently, 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) emerged as a promising approach, achieving both fast training and rendering time by representing the 3D scene with a large number of 3D Gaussians. This approach is optimized

^{*}Equal contribution ¹Interdisciplinary Program of Artificial Intelligence, Seoul National University, Seoul, South Korea ²School of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea ³School of Aerospace and Mechanical Engineering, Seoul National University, Seoul, South Korea. Correspondence to: Jonghyun Choi <jonghyunchoi@snu.ac.kr>.

by minimizing the reprojection error of the 3D Gaussians, while adding or removing them dynamically through the Adaptive Density Control (ADC) step.

Despite its advantages, 3DGS has significant limitations, particularly due to ADC's reliance on heuristics, which often leads to suboptimal performance. To address these issues, several studies (Deng et al., 2024; Bulò et al., 2024; Ye et al., 2024) have been proposed to reduce heuristic dependency and improve rendering quality and robustness. Among these works, 3D Gaussian Splatting as Markov Chain Monte Carlo (3DGS-MCMC) (Kheradmand et al., 2024) effectively addresses some of the key limitations present in earlier approaches, demonstrating state-of-the-art performance.

However, 3DGS-MCMC still struggles to generate clean novel view images for forward-facing scenes, particularly when using random initialization, as illustrated in Fig. 4. This is a critical limitation, since a number of real-world applications of NVS, such as augmented/virtual reality, autonomous driving, and human face capturing, involve forward facing camera settings. The limited coverage and viewing angles of forward-facing scenes lead to challenges, such as poor depth estimation and sparse reconstructions, which existing methods fail to address effectively.

In this paper, we propose Forward Facing 3D Gaussian Splatting as Markov Chain Monte Carlo (FF-3DGS-MCMC), a forward-facing NVS method based on 3DGS-MCMC. Our approach leverages depth map initialization to effectively mitigate the limitations of incomplete scene coverage in forward-facing scenarios, while refining the heuristic components of 3DGS. Additionally, we introduce a regularization scheme to prevent the Gaussians from overfitting to the closer regions to the camera in forward-facing setups. Together, these methods improve the robustness and quality of novel view synthesis in challenging scenarios.

2. Related Work

2.1. 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) is a method for representing 3D scenes by distributing a dense field of 3D Gaussians across the volume. Each 3D Gaussian is expressed as follows, where x is the position, Σ is the

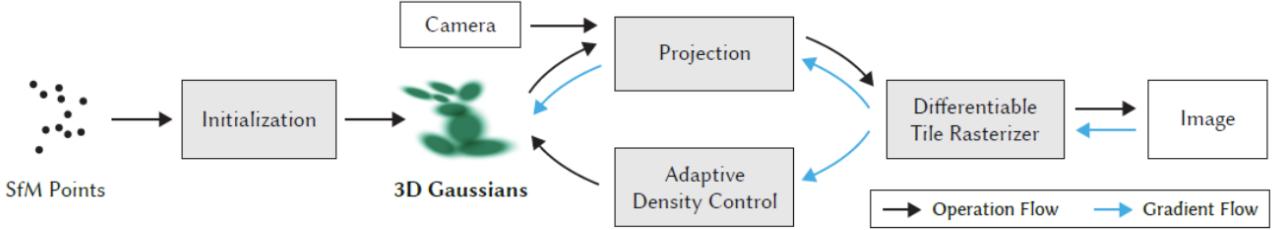


Figure 1. An overview of the 3D Gaussian Splatting pipeline. The Gaussians are initialized from Structure from Motion, then optimized through reducing projection error and Adaptive Density Control.

covariance matrix, and μ is the center of the Gaussian :

$$G(x|\mu, \Sigma) = e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (1)$$

The Gaussians also have parameters c and α , where they represent the color and opacity, respectively.

For rendering, the Gaussians are first projected onto the camera plane as an affine approximation, G^{2D} . The color of each pixel is calculated by the combination of projected 2D Gaussians by alpha blending:

$$C(x) = \sum_{i \in N} c_i \alpha_i G^{2D}(\mathbf{x}) \prod_{j=1}^{i-1} (1 - \alpha_j G^{2D}(\mathbf{x})), \quad (2)$$

where c_i is the color of each point and N is the amount of ordered points. Through the optimizing process, the Gaussians are altered by Adaptive Density Control (ADC), where redundant Gaussians are deleted, and essential Gaussians are cloned/split for handling poorly reconstructed regions and floating Gaussians that cause blurry reconstructions. The overview pipeline of 3DGS is visualized in Fig. 1.

2.2. 3DGS as Markov Chain Monte Carlo

While 3DGS has shown a remarkable ability for neural rendering, they still heavily rely on initialization strategies based on heuristics. To address these issues, (Kheradmand et al., 2024) proposed to reinterpret the placement and optimization of 3D Gaussians as a Markov Chain Monte Carlo (MCMC) sampling process, which eliminates the need for heuristic as well as human efforts.

It starts from the Stochastic Gradient Langevin Dynamics (SGLD), which is an MCMC-based approach that has recently been utilized in NVS (Brosse et al., 2018; Welling & Teh, 2011). Gradient g update for an image \mathbf{I} sampled from the set of training images \mathcal{I} by combined conventional 3DGS optimization and SGLD is as follows:

$$\mathbf{g} \leftarrow \mathbf{g} - \lambda_{lr} \cdot \nabla_{\mathbf{g}} \mathbb{E}_{\mathbf{I} \sim \mathcal{I}} [\mathcal{L}_{\text{total}}(\mathbf{g}; \mathbf{I})] + \lambda_{\text{noise}} \cdot \epsilon, \quad (3)$$

where λ_{lr} and λ_{noise} are the hyperparameters, and ϵ is the noise distribution.

In 3DGS-MCMC, relocating Gaussians to different positions can be viewed as a state transition to a new sample in the MCMC framework. Specifically, by preserving the sample probability within the MCMC process, it can be regarded as transitioning to another sample of equal probability, i.e., $\mathcal{P}(g^{\text{new}}) = \mathcal{P}(g^{\text{old}})$. In this manner, moving 'dead' Gaussians with low opacity ($\alpha_i < 0.005$) to the position of 'live' Gaussians with higher opacity reduces the reliance on heuristics and manual engineering, while maintaining stable rendering quality.

2.3. COLMAP-Free 3DGS

The 3D scene reconstruction and camera registration has been a challenging task in computer vision. COLMAP (Schönberger et al., 2016) is the Structure-from-Motion (SfM) library used to calculate camera positions and create a basic 3D point cloud from multiple images. Despite its convenience, this preprocessing step requires long processing time and lacks differentiability. Recent works (Wang et al., 2021; Lin et al., 2021; Bian et al., 2023) have focused on reducing reliance on SfM by simultaneously optimizing camera poses and the NeRF framework.

In traditional 3D Gaussian Splatting, COLMAP output helps initialize the Gaussians in the right spots in 3D space. COLMAP-Free 3D Gaussian Splatting (CF-3DGS) (Fu et al., 2024) replaces this with a neural network that directly estimates both the 3D geometry and camera positions from the images alone. The algorithm processes input frames sequentially and grows the 3D Gaussians set progressively by taking one input frame at a time, without pre-computing the camera poses. This makes the process faster and simpler, while still delivering high-quality results. The pipeline of CF-3DGS is visualized in Fig. 2.

2.4. Correspondence-Guided SfM-Free 3DGS

Although CF-3DGS (Fu et al., 2024) enables optimization of 3DGS (Kerbl et al., 2023) on multi-view images without relying on Structure-from-Motion (Schönberger & Frahm, 2016), a major drawback of this approach is the weak con-

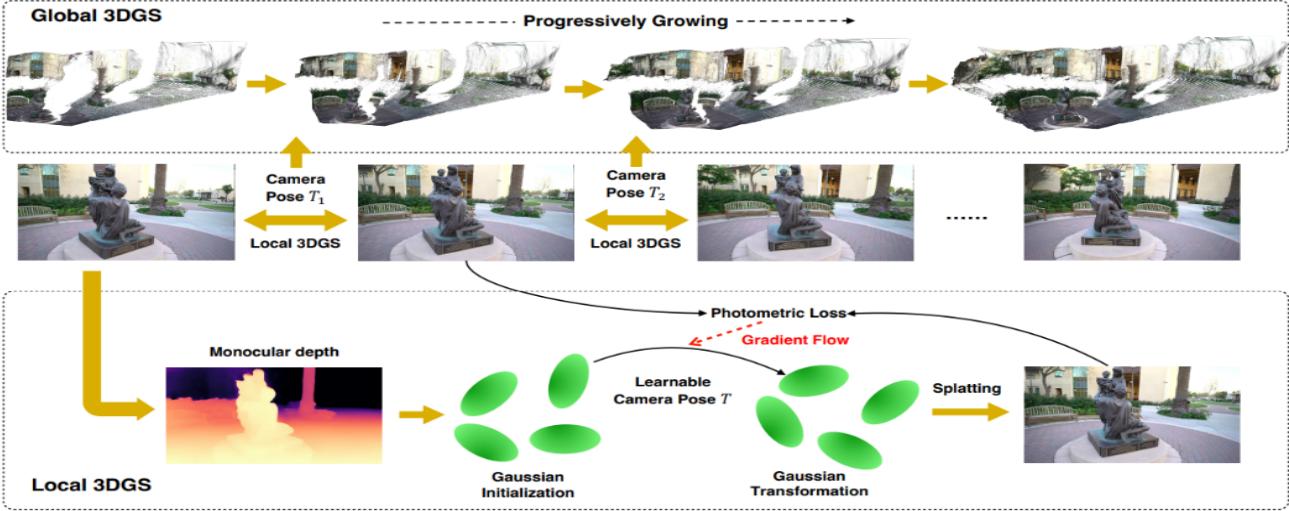


Figure 2. The pipeline of COLMAP-Free 3D Gaussian Splatting. The algorithm processes the input sequentially, achieving both camera pose estimation and 3D scene reconstruction.

straints leading to the suboptimal solutions. Specifically, CF-3DGS minimizes the L2 difference between observations and renderings, which is a well-known non-convex problem (Hartley & Zisserman, 2003) that complicates finding the optimal solution as highlighted by iNeRF (Yen-Chen et al., 2021).

Recently, Correspondence-Guided SfM-Free 3DGS (CG-3DGS) (Sun et al., 2024) introduced a correspondence-based loss to address the under-constrained optimization problem by adding stronger geometric conditions. Motivated by the powerfulness of multi-view geometry (Hartley & Zisserman, 2003), CG-3DGS presents the correspondence-based loss which enforces the alignment of features extracted from two renderings, adjusting the relative pose between two images with unknown initial poses. Consequently, (Sun et al., 2024) surpasses the traditional known-pose settings using SfM initialization. However, similar to CF-3DGS, CG-3DGS is limited to continuous observations (*i.e.* video frames), which restricts its applicability.

2.5. InstantSplat

DUSt3R (Wang et al., 2024) is one of the transformer based approaches for 3D scene reconstruction which uses a 2D-to-3D mapping network. Trained on a large scale dataset of 8.5M image and point cloud pairs, DUSt3R takes a pair of input images and outputs the 3D location of each pixel and the confidence of its reconstruction.

InstantSplat (Fan et al., 2024) is a Gaussian Splatting variation that uses DUSt3R as the backbone network. This method uses the estimated camera pose and point cloud obtained by the network for initializing 3DGS. Since DUSt3R

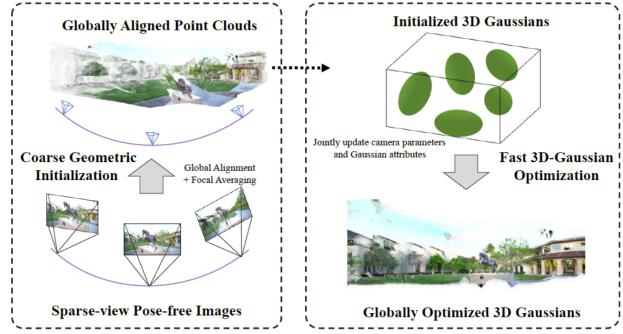


Figure 3. An overview of the InstantSplat pipeline. The algorithm uses DUSt3R to estimate the camera poses, then optimizes the Gaussians without Adaptive Density Control.

offers a much precise initial point cloud compared to SfM, InstantSplat does not go through the adaptive density control process. It optimizes the scene in only 1K iterations, which is a significant reduction compared to the 30K iteration of 3DGS (Kerbl et al., 2023).

2.6. DepthSplat

DepthSplat (Xu et al., 2024) merges 3D Gaussian splatting with depth estimation, demonstrating how these two tasks can enhance each other. It introduces a multi-view depth model that combines pre-trained monocular depth features from Depth Anything V2 (Yang et al., 2024) with multi-view feature matching. The predicted depths from the model are used as Gaussian centers for 3DGS, enabling high-quality novel view synthesis without per-scene optimization.



Figure 4. Novel View Synthesis result of 3DGS-MCMC for a forward-facing dataset (Left) and the ground truth point cloud (Right). The reconstruction contains floating artifacts near the rendering camera, significantly degrading the reconstruction quality.

DepthSplat also provides a method for pre-training the depth model in a fully unsupervised manner. Exploiting only the Gaussian Splatting rendering loss, which is the reprojection error of the generated novel view from 3DGS, the pre-training process shows significant enhancements especially for challenging datasets.

3. Approach

Although 3DGS-MCMC (Kheradmand et al., 2024) successfully reduces heuristic dependencies in the Adaptive Density Control process while achieving superior rendering performance, it encounters challenges in forward-facing data setups, as illustrated in Fig. 4. In order to address the limitations of the current 3DGS-MCMC algorithm, we propose Depth Supervision which provides extra geometrical information, with Near Gaussian Regularizers that prevent overfitting in regions close to the camera.

3.1. Depth Supervision

3DGS-MCMC determines the optimal locations for Gaussian components through random exploration. When applied to 360-degree scenes represented by MipNeRF360 dataset (Barron et al., 2022), the possible error from random walking Gaussians are typically corrected by observations from other viewing directions. However, forward-facing scenes (Mildenhall et al., 2019) present a unique challenge since all cameras are oriented in a similar direction, where any random walk along the scene’s z-axis can still appear valid, resulting in inherent ambiguity in the z-direction as shown in Fig. 4. To address z-direction ambiguity, we incorporate an additional depth information by incorporating an off-the-shelf depth estimation network.

We employ ZoeDepth (Bhat et al., 2023) to provide the depth guidance. Given N multi-view observations $\{I_i\}_{i=1}^M$, we obtain M estimated depth maps $D_i = F_\theta(I_i)$, where θ is a pre-trained ZoeDepth parameter. Meanwhile, we acquire a rendered depth from given camera poses by alpha blending (Eq. 2) of z-coordinate instead of colors, inspired



Figure 5. Overfitting of near-located Gaussian visualized by rendering RGB (Left) and depth (Right). Gaussians located near the camera screen becomes suboptimal solution of 3DGS-MCMC (Kheradmand et al., 2024) even with the depth loss.

by previous works (Zhu et al., 2024; Chung et al., 2023):

$$d(\mathbf{x}) = \sum_{i \in N} z_i \alpha_i G^{2D}(\mathbf{x}) \prod_{j=1}^{i-1} (1 - \alpha_j G^{2D}(\mathbf{x})), \quad (4)$$

where z_i is a z-coordinate of corresponding Gaussian in the camera coordinate. The rendered depth map is defined as $\hat{D}_i = \{d(\mathbf{x}) | x \in \mathcal{P}_i\}$ where \mathcal{P}_i stands for all pixels for i -th image. We adopt the depth-guide loss to guide the rendered depth \hat{D}_i toward the estimated depth D_i invariant to scale and shift:

$$\mathcal{L}_{\text{depth}} = \frac{\text{Cov}(D_i, \hat{D}_i)}{\sqrt{\text{Var}(D_i)\text{Var}(\hat{D}_i)}}. \quad (5)$$

Since the predicted depth from ZoeDepth is relative depth which is normalized to $[0, 1]$, $\mathcal{L}_{\text{depth}}$ guides the distribution difference of the depth prediction instead of L2-distance, similar to (Zhu et al., 2024).

3.2. Regularizing Near Gaussian

Although $\mathcal{L}_{\text{depth}}$ restricts overly free exploration along the z-axis in forward-facing scenes, there still exists a room for ambiguity as $\mathcal{L}_{\text{depth}}$ only guides the relative distribution of the depth map. A common suboptimal solution from this ambiguity is demonstrated in Fig. 5, occurring when a Gaussian attaches itself directly in front of a camera screen (*i.e.* a near plane). To address the Gaussians located near the camera screen, we introduce a near regularizer that reduces the opacity of Gaussians positioned too close to the camera. Specifically, we penalize the Gaussian opacity whose depth is under certain threshold. This adjustment allows the gaussian exploration depicted in Eq. 3 to transfer Gaussians at the near plane to the optimal states by state transition. We penalized the influence of Gaussians with depths smaller than a hyperparameter z_{pivot} on the rendering process, proportional to their impact on a logarithmic scale:

$$\mathcal{L}_{\text{near}} = \sum_i w_i \max(\log(z_{\text{pivot}}) - \log(z_i), 0), \quad (6)$$

where $w_i = G_i^{2D}(\mathbf{x})\alpha_i$ is an alpha-blending weight. We empirically set $z_{\text{pivot}} = 0.05z_{\text{far}}$ where z_{far} is a pre-defined



(a) Rendered images with 3DGS-MCMC



(b) Rendered images with FF-3DGS-MCMC (Ours)

Figure 6. Qualitative comparison of rendered images. (a) Rendered images generated using 3DGS-MCMC method, demonstrating the baseline approach’s visual quality, and (b) Rendered images from our method, FF-3DGS-MCMC, highlighting improvements in visual fidelity and structural consistency compared to (a).

furthest distance from the camera (Kerbl et al., 2023).

3.3. Training Loss

We finalize our training loss \mathcal{L} by:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda_{\text{depth}} \mathcal{L}_{\text{depth}} + \lambda_{\text{near}} \mathcal{L}_{\text{near}}, \quad (7)$$

where \mathcal{L}_{rec} is the reconstruction loss from vanilla 3DGS (Kerbl et al., 2023). We choose the scalar $\lambda_{\text{depth}} = 0.01$ and $\lambda_{\text{near}} = 0.001$.

4. Evaluation

4.1. Experimental Setup

Dataset We use LLFF dataset (Mildenhall et al., 2019), which is commonly used in vanilla NeRF. This dataset includes eight scenes such as fern, flower, and fortress, with each scene comprising approximately 15 to 60 multi-view images in a forward-facing settings. We conduct our experiment using images with a resolution of 1024×756 , which corresponds to $\times 4$ downscaling from the original images, similar to the setup in NeRF. For the test views, we excluded images with indices that are multiples of 8 from training to ensure a consistent evaluation of the model’s performance. Camera poses are obtained using COLMAP, and

the point cloud is randomly initialized by first determining the mean pose and then uniformly distributing a total of 100,000 points within its frustum.

Performance metrics We use two types of metrics to evaluate image quality: Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Metric (SSIM). Higher PSNR, and higher SSIM values indicate better model performance.

Baselines in comparison For performance comparison in §4.2, we compare the performance of FF-3DGS-MCMC with existing Gaussian splatting technique, 3DGS-MCMC. We validate our proposed method shows comparable results, as discussed in §1.

4.2. Performance comparison

In this subsection, we present a performance evaluation of our FF-3DGS-MCMC method in comparison with the baseline method, 3DGS-MCMC, as introduced in §4.1. Implicit methods (Mildenhall et al., 2020; Barron et al., 2022) have been

Qualitative results Fig. 6 illustrates a qualitative comparison of rendered images. In Fig. 6 (a), images generated by the 3DGS-MCMC method are presented, showing float-

Method	Fern		Flower		Fortress		Horns	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
3DGs-MCMC (Kheradmand et al., 2024)	17.62	0.613	18.12	0.618	29.16	0.886	25.57	0.840
FF-MCMC (Ours)	23.02	0.765	25.05	0.785	29.16	0.862	25.78	0.827
Method	Leaves		Orchids		Room		Trex	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
3DGs-MCMC (Kheradmand et al., 2024)	9.68	0.162	13.69	0.392	32.48	0.960	28.43	0.932
FF-MCMC (Ours)	9.37	0.082	19.84	0.679	29.04	0.930	26.36	0.889
						Average		
							PSNR↑	SSIM↑

Table 1. Quantitative comparison of novel view synthesis on the LLFF Dataset.

Method	PSNR↑	SSIM↑
Baseline	21.84	0.675
$\mathcal{L}_{\text{depth}}$ only	23.19	0.727
$\mathcal{L}_{\text{near}}$ only	23.35	0.730
Ours	23.45	0.727

Table 2. Ablation Study.

ing artifacts. Fig. 6 (b) showcases the images rendered using our FF-3DGs-MCMC method, which show significant improvements in visual fidelity and structural consistency. Specifically, our method effectively preserves fine details and reduces the number of blurry artifacts, leading to higher perceptual quality. This aligns with the quantitative results discussed in the following section.

Quantitative results The quantitative evaluation of our FF-3DGs-MCMC method compared to the baseline 3DGs-MCMC is presented in Table 1. Notably, for some objects, the performance difference is particularly significant. This can be attributed to the sensitivity of the rendering quality to the number of Gaussian points used. In our evaluation, we generated a default of 100,000 Gaussians for all objects to ensure consistency across comparisons. This highlights the robustness of our method under standard conditions while addressing the limitations observed in the baseline approach.

4.3. Ablation Study

We present an ablation study on the LLFF dataset (Mildenhall et al., 2019) to evaluate and analyze the impact of each component. As shown in Table 2, every component contributes significantly to the novel view synthesis task, with the best performance achieved when all elements are integrated. Using either depth supervision or the near regularizer alone can significantly reduce floating artifacts such as Fig 3.1, leading to substantial performance improvements with one additional loss term. However, as demonstrated in Fig 5, a single loss cannot address all types of artifact and the best performance is achieved when both depth supervision and the near regularizer are applied together.

5. Discussion

Extension to dynamic scenes Integrating dynamic scenes represents a natural and promising extension, but introduces challenges that go beyond the main purpose of this paper. Dynamic scenes typically require the use of 4D Gaussian splatting, incorporating temporal information or deformation fields, which significantly increases the complexity of the approach. Furthermore, ADC in dynamic scenarios poses additional challenges distinct from those addressed in static settings. Although it goes beyond the current scope, this direction holds great potential for future exploration.

Benchmark extension In this work, we primarily focus on evaluating our method against other explicit reconstruction approaches, specifically 3DGs-MCMC. One of the widely used methods in NVS, NeRF is considered implicit reconstruction method, rather than explicitly defining geometry or surfaces, making a direct comparison challenging. This approach of separating implicit and explicit methods for comparison has also been conducted in previous research (Huang et al., 2024).

6. Conclusion

In this work, we propose Forward Facing 3D Gaussian Splatting as Markov Chain Monte Carlo, an enhanced approach to 3D Gaussian Splatting as Markov Chain Monte Carlo (3DGs-MCMC) to address its limitations in forward-facing scenes. While 3DGs-MCMC has demonstrated state-of-the-art performance in Novel View Synthesis(NVS), it faces challenges in maintaining high quality results for forward-facing scenarios, which are critical in applications such as augmented/virtual reality, autonomous driving, and human face capturing. By incorporating depth maps into the optimization process, our method improves the representation of forward-facing scenes, leading to more accurate and visually consistent novel view synthesis results. The integration of depth information enhances robustness across various forward-facing scene configurations and reduces heuristics dependencies observed in existing methods. Experimental results on the LLFF dataset demonstrate that our method outperforms the baseline, particularly in challeng-

ing forward-facing scenes. These findings underline the effectiveness of our approach in advancing NVS and its potential for real-world applications, paving the way for future research in depth-guided 3D Gaussian Splatting techniques.

References

- Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., and Hedman, P. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5470–5479, 2022.
- Bhat, S. F., Birk, R., Wofk, D., Wonka, P., and Müller, M. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*, 2023.
- Bian, W., Wang, Z., Li, K., Bian, J., and Prisacariu, V. A. Nope-nerf: Optimising neural radiance field with no pose prior. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2023.
- Brosse, N., Durmus, A., and Moulines, E. The promises and pitfalls of stochastic gradient langevin dynamics. *Advances in Neural Information Processing Systems*, 31, 2018.
- Bulò, S. R., Porzi, L., and Kortschieder, P. Revising densification in gaussian splatting. *arXiv preprint arXiv:2404.06109*, 2024.
- Chung, J., Oh, J., and Lee, K. M. Depth-regularized optimization for 3d gaussian splatting in few-shot images. *arXiv preprint arXiv:2311.13398*, 2023.
- Deng, X., Diao, C., Li, M., Yu, R., and Xu, D. Efficient density control for 3d gaussian splatting. *arXiv preprint arXiv:2411.10133*, 2024.
- Fan, Z., Cong, W., Wen, K., Wang, K., Zhang, J., Ding, X., Xu, D., Ivanovic, B., Pavone, M., Pavlakos, G., et al. Instantsplat: Unbounded sparse-view pose-free gaussian splatting in 40 seconds. *arXiv preprint arXiv:2403.20309*, 2, 2024.
- Fu, Y., Liu, S., Kulkarni, A., Kautz, J., Efros, A. A., and Wang, X. Colmap-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 20796–20805, June 2024.
- Hartley, R. and Zisserman, A. *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- Huang, B., Yu, Z., Chen, A., Geiger, A., and Gao, S. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 Conference Papers*, pp. 1–11, 2024.
- Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.
- Kheradmand, S., Rebain, D., Sharma, G., Sun, W., Tseng, J., Isack, H., Kar, A., Tagliasacchi, A., and Yi, K. M. 3d gaussian splatting as markov chain monte carlo. *arXiv preprint arXiv:2404.09591*, 2024.
- Lin, C.-H., Ma, W.-C., Torralba, A., and Lucey, S. Barf: Bundle-adjusting neural radiance fields. In *International Conference on Computer Vision*, 2021.
- Mildenhall, B., Srinivasan, P. P., Ortiz-Cayon, R., Kalantari, N. K., Ramamoorthi, R., Ng, R., and Kar, A. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics*, 38(4):1–14, 2019.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *European Conference on Computer Vision*, pp. 405–421, 2020.
- Schönberger, J. L. and Frahm, J.-M. Structure-from-motion revisited. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2016.
- Schönberger, J. L., Zheng, E., Pollefeys, M., and Frahm, J.-M. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision*, 2016.
- Sun, W., Zhang, X., Wan, F., Zhou, Y., Li, Y., Ye, Q., and Jiao, J. Correspondence-guided sfm-free 3d gaussian splatting for nvs. *arXiv preprint arXiv:2408.08723*, 2024.
- Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., and Revaud, J. Dust3r: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 20697–20709, 2024.
- Wang, Z., Wu, S., Xie, W., Chen, M., and Prisacariu, V. A. NeRF—: Neural radiance fields without known camera parameters. *arXiv preprint arXiv:2102.07064*, 2021.
- Welling, M. and Teh, Y. W. Bayesian learning via stochastic gradient langevin dynamics. In *International Conference on Machine Learning*, pp. 681–688. Citeseer, 2011.
- Xu, H., Peng, S., Wang, F., Blum, H., Barath, D., Geiger, A., and Pollefeys, M. Depthsplat: Connecting gaussian splatting and depth. *arXiv preprint arXiv:2410.13862*, 2024.
- Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., and Zhao, H. Depth anything v2. *arXiv preprint arXiv:2406.09414*, 2024.

Ye, Z., Li, W., Liu, S., Qiao, P., and Dou, Y. Absgs: Recovering fine details in 3d gaussian splatting. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 1053–1061, 2024.

Yen-Chen, L., Florence, P., Barron, J. T., Rodriguez, A., Isola, P., and Lin, T.-Y. iNeRF: Inverting neural radiance fields for pose estimation. In *International Conference on Intelligent Robots and Systems*, 2021.

Zhu, Z., Fan, Z., Jiang, Y., and Wang, Z. Fsgs: Real-time few-shot view synthesis using gaussian splatting. In *European Conference on Computer Vision*, pp. 145–163. Springer, 2024.