The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W6-2025
3rd International Workshop on Evaluation and BENCHmarking of Sensors, Systems and
GEOspatial Data in Photogrammetry and Remote Sensing (GEOBENCH), 20–21 November 2025, Wroclaw, Poland

Evaluating 3D Gaussian Splatting for Urban Scene Reconstruction

Ziyang Yan [1,2], Mengrui Yin [3], Yihua Shao [4], Fabio Remondino [1]

[1]3D Optical Metrology Unit, Bruno Kessler Foundation (FBK), Trento, Italy – Email: (zyan, remondino)@fbk.eu
[2]DISI, University of Trento, Trento, Italy
[3] College of Resources and Environment, Chengdu University of Information Technology, Chengdu, China
- Email: mengruiyin6@gmail.com
[4]Department of Computer Science, City University of Hong Kong, Hong Kong – Email: yihuajerry@gmail.com

Keywords: Urban scene reconstruction, General evaluation, 3D Gaussian Splatting.

Abstract

Accurate, detailed and efficient 3D reconstructions of large-scale urban environments are essential for applications such as autonomous driving, urban planning and digital twin construction. Recent advances in 3D Gaussian Splatting (3DGS) have shown remarkable potential in photorealistic novel view synthesis and high-fidelity scene reconstruction, but their applicability to large-scale urban reconstruction remains underexplored and often challenging. In this work, we present a comprehensive evaluation of 3D Gaussian Splatting techniques applied to urban scale 3D reconstruction. We systematically benchmark GS-based methods on diverse urban datasets, analyzing their performance in terms of scalability, geometric accuracy, rendering quality and computational efficiency. The study aims to bridge the gap between emerging 3DGS research and real-world urban reconstruction requirements, offering insights and guidelines for deploying Gaussian Splatting in practical large-scale scenarios.

1 Introduction

Urban-scale 3D scene reconstruction is a fundamental problem in computer vision, photogrammetry, and graphics, serving as the backbone for applications such as autonomous driving [Yan et al., 2025a, Wang et al., 2025a, Shao et al., 2025a, Shao et al., 2024b], urban planning [Zhang et al., 2021], and digital twins [Dong et al., 2025, Roman et al., 2025]. The core objective lies in achieving photorealistic rendering alongside accurate and complete geometric reconstruction, often under strict efficiency constraints. While Neural Radiance Fields (NeRF) [Mildenhall et al., 2021] has driven substantial progress in novel view synthesis, 3D Gaussian Splatting (3DGS) [Kerbl et al., 2023] has recently emerged as a dominant paradigm due to its fast training convergence, high rendering efficiency, and competitive visual fidelity. By representing a scene as a set of discrete Gaussian ellipsoids and rendering via a highly optimized rasterization pipeline, 3DGS offers a promising balance between quality and speed [Yan et al., 2024]. However, when extending 3DGS from object-level or small-scale scenes to complex urban environments, a number of significant challenges arise. First, the unordered Gaussian representation lacks explicit surface correspondence, which can degrade synthesis quality in extrapolated viewpoints and limit downstream usability in tasks such as semantic editing, dynamic object integration, and realistic relighting [Guédon and Lepetit, 2024]. Second, scaling to large urban datasets exacerbates issues related to memory usage [Akhavi Zadegan et al., 2025], computational overhead, and geometric fidelity. For example, mesh-coupled approaches like SuGaR [Guédon and Lepetit, 2024] can improve surface consistency but often fail to capture fine urban structural details and suffer from severe memory overheads. Similarly, Gaussian Opacity Fields [Yu et al., 2024] can produce oversized, blurred Gaussians that block visibility and cause underfitting, while 2DGS [Huang et al., 2024] faces convergence slowdowns and Gaussian count explosions in parallel training scenarios. In addition to representation and scalability issues, evaluation itself poses unique difficulties for urban scenes [Liao et al., 2022]. Sparse or occluded boundary regions—common in street-level and aerial captures—lead to unstable geometry estimation and high metric variance , complicating fair comparison between methods. Moreover, practical deployment demands efficient training and compression strategies: large-scale 3DGS reconstructions can require tens of millions of Gaussians, consuming over 30 GB of GPU memory and reducing rendering speeds below real-time. While methods like VastGaussian [Lin et al., 2024], CityGaussian [Liu et al., 2024a] and CityGaussianv2 [Liu et al., 2024b] address scalability, their multi-hour training pipelines remain prohibitive for time-critical or resource-constrained settings. These limitations highlight the need for a systematic evaluation of 3DGS-based reconstruction in urban-scale environments. In this work, we present a comprehensive benchmarking of state-of-the-art 3DGS variants on diverse urban datasets (Table 1), analyzing their performance in terms of geometric accuracy, rendering quality, scalability, efficiency, and robustness to challenging urban conditions. Our study aims to provide actionable insights for both researchers and practitioners, bridging the gap between emerging 3DGS research and the demands of real-world urban reconstruction.

2 Related Works

2.1 Large Scale Scene Reconstruction

Large-scale scene reconstruction has evolved from traditional geometry-based pipelines to modern neural representations. Early methods relied on SfM [Snavely et al., 2006] and MVS [Seitz et al., 2006], with systems such as PMVS [Furukawa and Ponce, 2009], Building Rome in a Day [Agarwal et al., 2011], and COLMAP demonstrating city-scale reconstruction from Internet photo collections. Later, learning-based MVS approaches, including MVSNet [Yao et al., 2018], Fast-MVSNet [Yu and Gao, 2020], and PatchmatchNet [Wang et al., 2021], improved robustness and scalability. The introduction of NeRF [Mildenhall et al., 2021] marked a paradigm shift, achieving high-quality novel view synthesis but facing scalability challenges [Remondino et al., 2023]. To overcome this, works such as Block-NeRF [Tancik et al., 2022], Mega-NeRF [Turki et al., 2022], Urban Radiance Fields [Rematas et al., 2022], and Zip-NeRF [Barron et al., 2023] extended NeRF to urban and city-scale scenes. Most recently, 3D Gaussian Splatting (3DGS) [Kerbl et al., 2023] introduced an

Dataset	Sample Image	Numb.of images	Camera type & image size	Description / GDS	Ground Truth (GT)
UseGeo - Dataset_1[1]		224	Sony ILCE-7RM3, 7952x5304 px	High-resolution built-up and vegetated areas (1.7 cm GSD)	Helicopter-based Laser Scanner
Dortmund[2]		59	IGI PentaCam, 6132x8176 px (N), 8176x6132 px (O)	Outdoor large scale, built-up and vegetated areas (12 cm GSD)	Airborne-based Laser Scanner
MatrixCity Block_2[3]		535	Virtual camera 1920x1080 px (synthetic images)	Urban large-scale, metropolitan areas (6.5cm GSD)	MVS / Photogrammetry

[1] https://github.com/3DOM-FBK/usegeo
[2] https://github.com/3DOM-FBK/NeRFBK
[3] https://city-super.github.io/matrixcity/

Table 1. The large-scale urban scene datasets used for evaluating Gaussian Splatting methods.

explicit and efficient representation, enabling real-time rendering and opening new directions for large-scale reconstruction [Shao et al., 2025b, Shao et al., 2025d]. In the next subsection, we focus on 3DGS approaches tailored to large-scale settings.

2.2 3D Gaussian Splatting

Since the introduction of 3D Gaussian Splatting (3DGS), numerous attempts tried to adopt it to large-scale urban reconstruction. BlockGaussian [Wu et al., 2025] partitions urban environments into smaller spatial blocks, enabling distributed training and rendering, while CityGaussian [Liu et al., 2024a] scales 3DGS to city-level datasets but suffers from Gaussian redundancy. CityGaussianv2 [Liu et al., 2024b] addresses scalability with gradient-decomposed densification, depth regression, and elongation filtering, improving convergence and reducing memory usage. Wild-GS [Xu et al., 2024] targets in-the-wild imagery with diverse viewpoints, though its stability in structured urban scenes is limited. Ostree-GS [Ren et al., 2024] employs octree-based subdivision for hierarchical Gaussian management, achieving better memory efficiency at the expense of fine-grained detail. Despite these advances, most methods rely on SfM initialization and struggle in low-texture or distant regions [Liao et al., 2025]. To overcome such limitations, hybrid or enhanced representations have been proposed. HO-Gaussian [Li et al., 2024] integrates grid-based volumes and introduces Gaussian Direction Encoding for robust urban rendering, while HUGS [Zhou et al., 2024] extends 3DGS to holistic scene understanding via joint optimization of geometry, semantics, and motion. Scalability and efficiency are further improved by semantic or hierarchical partitioning [Shao et al., 2025c, Shao et al., 2024a]: PG-SAG [Wang et al., 2025b] leverages semantic-aware grouping for fine-grained building reconstruction, CityGS-X [Gao et al., 2025]introduces a parallelized hybrid hierarchical structure for multi-GPU training, and GigaGS [Chen et al., 2024] applies visibility-based partitioning with photometric and geometric consistency for large-scale surface reconstruction. In addition, FlashGS [Feng et al., 2025] optimizes algorithms and system pipelines to achieve an order-of-magnitude speedup, and Horizon-GS [Jiang et al., 2025] unifies aerial and ground views for immersive city-scale navigation. Collectively, these works demonstrate the growing potential of 3DGS for urban-scale applications, while highlighting ongoing challenges in balancing scalability, geometric fidelity, and multi-view consistency.

2.3 Benchmark for Urban Scene Reconstruction

While recent advances in learning-based approaches have significantly improved their scalability and efficiency in case of large-scale urban reconstruction, their evaluation still relies on high-quality datasets that capture diverse urban settings. To fairly benchmark 3D urban reconstruction methods and enable consistent comparisons, several synthetic and real-world datasets have been curated. The most recent and widely used benchmarks / datasets are:

- UrbanScene3D [Lin et al., 2022]: UrbanScene3D is a comprehensive dataset for urban scene perception and reconstruction, featuring over 128,000 high-resolution images across 16 scenes, encompassing both real urban regions and synthetic cities over a total area of 136 km². It includes high-precision LiDAR scans, 3D meshes, depth maps, and semantic annotations, supporting research in aerial path planning, 3D reconstruction, and semantic segmentation.

- City3D [Huang et al., 2022]: City3D is a dataset focused on large-scale building reconstruction from airborne LiDAR point clouds. It contains point clouds and reconstructed 3D models of 20,000 real-world buildings. The dataset addresses challenges such as missing vertical walls in LiDAR data and provides a benchmark for evaluating building reconstruction algorithms and urban modeling techniques.

- Mill19 [Turki et al., 2022]: Mill19 is a large-scale industrial site dataset, including two scenes: buildings and ruins. It provides thousands of high-resolution images captured primarily by UAVs, covering over 100,000 m². The dataset is used to evaluate various of reconstruction methods in complex urban environments.

- MatrixCity [Li et al., 2023]: MatrixCity is a large-scale synthetic dataset designed for city-scale neural rendering research. It comprises 67,000 aerial images and 452,000 street-level images from two city maps totaling 28 km². The dataset offers diverse environmental conditions, including varying lighting and weather, and provides depth maps,

camera trajectories, and semantic annotations, facilitating the development and evaluation of neural rendering models for urban environments.

- NeRFBK [Yan et al., 2023]: NeRFBK includes two ultra-high-resolution urban-scale datasets from Dortmund, Germany. The city center dataset (UseGeo) consists of UAV imagery (338 nadir / oblique shots at 1–3 cm GSD) captured with a DJI S800. And the Dortmund dataset integrates a 5-camera oblique system along with 59 UAV captures. Both datasets provide comprehensive ground-truth for evaluating reconstruction quality.
- Horizon-GS [Jiang et al., 2025]: Horizon-GS provides 5 synthetic city-scale scenes and 2 real-world UAV + street-level captures. Synthetic data includes RGB, depth, and point clouds for 3D-GS training, while real data combines drone and helmet-mounted cameras with SfM-based reconstruction. Depth is refined with Depth-Anything-V2, and moving objects/sky are removed for cleaner supervision.

3 3DGS background

3D Gaussian Splatting is an explicit scene representation method that encodes geometry through parameterized primitives such as voxel-like grids or point-based Gaussians. The concept of Gaussian splatting has its roots in Elliptical Weighted Average (EWA) splatting, but its recent application to high-quality 3D reconstruction and novel view synthesis was pioneered by [Kerbl et al., 2023], establishing a major step forward in the field. The pipeline typically begins with an image orientation step and a sparse point cloud generation, which is used to initialize the set of 3D Gaussians. Each Gaussian point x is described by a center μ and a covariance matrix Σ in world coordinates. The Gaussian distribution is written as:

$$G(x) = \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right). \quad (1)$$

To ensure that Σ remains symmetric and positive definite, it is decomposed into a rotation matrix R and a scaling matrix S:

$$\Sigma = RSS^\top R^\top. \quad (2)$$

Projection onto the 2D image plane is performed by computing a transformed covariance Σ' in camera space:

$$\Sigma' = JW\Sigma W^\top J^\top, \quad (3)$$

where J is the Jacobian of the affine projection and W is the camera transformation matrix. Rendering is performed via alpha compositing, where the final pixel color is accumulated in depth order of Gaussians. The pixel intensity C is computed as:

$$C = \sum_{i=1}^{N} c_i \alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \quad (4)$$

with c_i the color of the i-th Gaussian and α_i its blending weight, derived from opacity and projected density. Unlike implicit neural fields such as NeRF, 3DGS offers real-time rendering by explicitly storing scene elements as Gaussians. The approach combines differentiable optimization with efficient GPU rasterization techniques, including depth sorting and tile-based rendering, which support anisotropic splats. Each Gaussian is parameterized by position, covariance, color, and opacity, and these values are optimized using stochastic gradient descent (SGD) [Ruder, 2016] within a fully differentiable rendering framework. The optimization process minimizes the difference between the rendered images and the ground-truth frames. A differentiable rasterizer propagates gradients from image space back to the Gaussian parameters. The objective combines pixel-wise $\mathcal{L}_1$ loss with a perceptual Structural Similarity Index (SSIM) term:

$$\mathcal{L} = (1-\lambda), \mathcal{L}1 + \lambda, \mathcal{L}\text{D-SSIM}, \quad (5)$$

where λ balances pixel fidelity with structural similarity. This formulation allows 3D Gaussian Splatting to deliver high-quality photorealistic results at interactive rendering speeds.

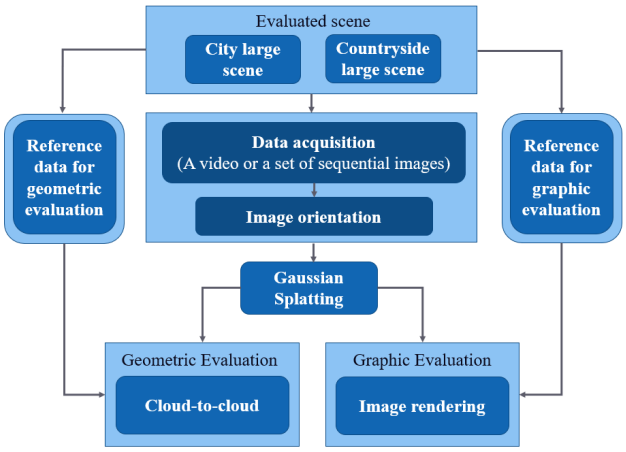


Figure 1. Overall framework for assessing the performance of various Gaussian Splatting methods in urban scene reconstruction

4 Methodology

4.1 Assessment methodology

We design our evaluation pipeline (shown in Figure 1) following the methodology presented in [Yan et al., 2025b]. Since 3DGS requires camera poses as input, we adopt the exterior parameters provided by each dataset, which were estimated using COLMAP. The selected Gaussian Splatting methods are then applied to generate dense 3D geometries. All generated point clouds are co-registered to the ground-truth (GT) reference using an Iterative Closest Point (ICP) algorithm in CloudCompare. Finally, qualitative and quantitative evaluations are performed based on established photogrammetric criteria to ensure unbiased assessment. Comprehensive results will be presented in the experimental section of the full paper.

4.2 Implementation details and data pre-processing

We implement all evaluations in PyTorch and CUDA, conducting experiments on a single NVIDIA Tesla H100 GPU. For

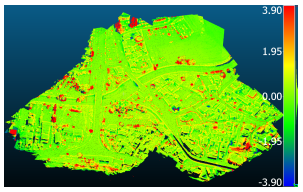
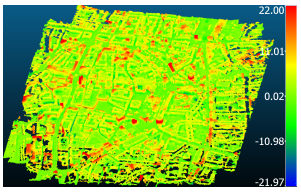
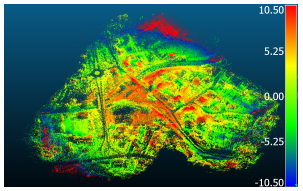
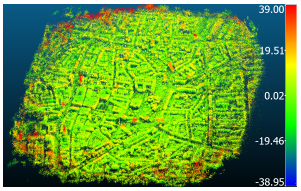
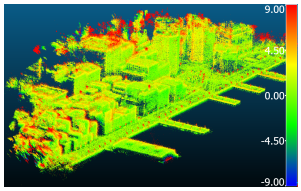
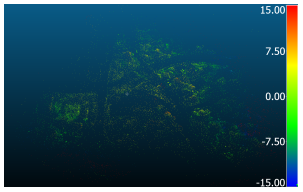
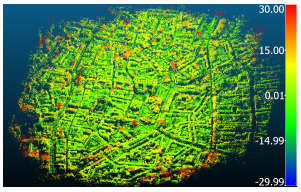
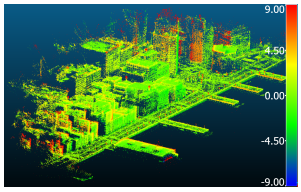
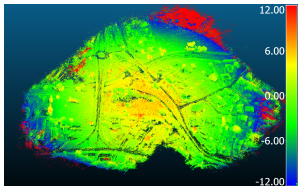
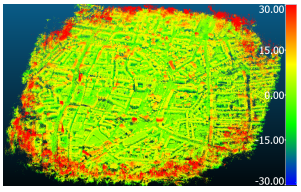
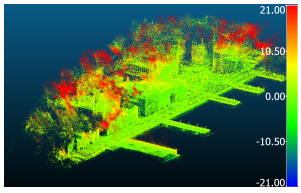
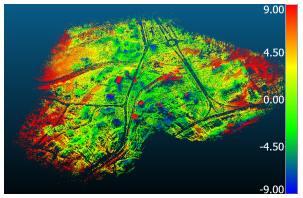
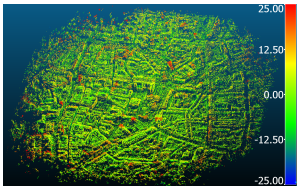
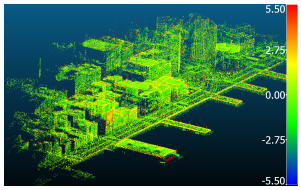
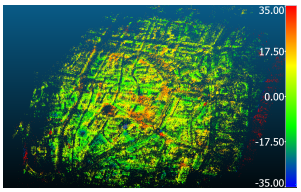
Method	UseGeo	Dortmund	MatrixCity
Photogrammetry			Not available
BlockGaussian [Wu et al., 2025]			
CityGaussian [Liu et al., 2024a]			
Gaussian Opacity Fields [Yu et al., 2024]			
Octree-GS [Ren et al., 2024]			
Wild-GS [Xu et al., 2024]	✗		✗

Table 2. Visual results of the geometry-based comparisons for the evaluated methods across all datasets. The legends in the figures are set to +/- 3σ [unit: m] considering the metrics of each methods (Table 3).

the five 3DGS approaches examined in this work, CityGaussian and BlockGaussian adopt a block-based training strategy, while Octree-GS, Gaussian Opacity Fields (GOF), and Wild-Gaussians employ a holistic training approach. Due to the high GPU memory consumption of block-based training, we downsample the image resolution to prevent out-of-memory errors and apply depth regularization (e.g., Depth Anything V2 [Yang et al., 2024]) to refine the sparse point clouds generated in COLMAP. This improves the quality of the initial models and compensates for the potential restruction loss caused by image downsampling. The remaining methods do not require special pre-processing; their main distinctions lie in training and Gaussian initialization. CityGaussian and BlockGaussian train in partitions for efficiency, Octree-GS enables adaptive multi-level optimization, and GOF and Wild-Gaussians enhance training with opacity learning and robustness to sparse data.

5 Experiment Results

Urban-scale datasets pose unique challenges for 3D reconstruction, especially due to complex building structures, large variations in scale, and reflective or textureless surfaces. We evaluated five Gaussian Splatting methods: BlockGaussian [Wu et al., 2025], Octree-GS [Ren et al., 2024], CityGaussian [Liu et al., 2024a], Gaussian Opacity Fields [Yu et al., 2024], and Wild-GS [Xu et al., 2024] on three representative urban datasets, namely Drone, Dortmund, and MatrixCity Block_2. We presents the qualitative results of derived 3D reconstructions across the three datasets in Table 2 while the quantitative comparisons of geometric and graphic are reported in Table 3.

5.1 Geometric Evaluation

5.1.1 Qualitative Evaluation of Geometric

As shown in Table 2, we subjectively annotate whether each reconstruction is deemed successful, defined as exhibiting no

Dataset	Method	RMSE ↓	MAE ↓	STD ↓	Mean_E ↓
UseGeo	Photogrammetry	0.23	0.18	0.14	0.20
	BlockGaussian	3.98	3.08	3.51	3.27
	Octree-GS	3.39	2.52	2.28	2.69
	CityGaussian	5.46	3.87	3.85	4.42
	Gaussian Opacity Fields	4.40	3.06	3.16	3.30
	Wild-GS	–	–	–	–
Dortmund	Photogrammetry	0.88	0.68	0.55	0.75
	BlockGaussian	3.50	2.71	2.48	2.92
	Octree-GS	1.63	1.07	1.23	1.29
	CityGaussian	1.77	1.11	1.38	1.48
	Gaussian Opacity Fields	3.14	2.13	2.32	2.68
	Wild-GS	6.85	4.97	4.71	5.75
MatrixCity	BlockGaussian	1.68	1.18	1.19	1.38
	Octree-GS	0.73	0.53	0.49	0.61
	CityGaussian	1.80	1.27	1.28	1.41
	Gaussian Opacity Fields	3.43	1.82	2.90	2.79
	Wild-GS	–	–	–	–

Table 3. Evaluation metrics [m] across different methods on the UseGeo, Dortmund and MatrixCity datasets. Best results are highlighted in green, worst results in red.

significant geometric discrepancies with respect to the GT. The qualitative results further indicate that Photogrammetry achieves the most accurate reconstructions, as most regions are correctly recovered (highlighted in green). In contrast, 3DGS methods tend to generate noisy points along scene boundaries (marked in red), which negatively affect their overall reconstruction performance.

5.1.2 Quantitative Evaluation

UseGeo - Dataset 1 It is particularly challenging due to its high-resolution nadir imagery (1.7 cm GSD) covering mixed urban and rural environments. Among all evaluated methods, Photogrammetry achieved the most accurate reconstruction, with RMSE = 0.23 m, MAE = 0.18 m, STD = 0.14 m, and Mean_E = 0.20 m, demonstrating its superior geometric fidelity. Octree-GS ranked second, achieving the best performance among 3DGS methods with RMSE = 3.39 m, MAE = 2.52 m, STD = 2.28 m, and Mean_E = 2.69 m. BlockGaussian and GOF produced acceptable results, whereas CityGaussian exhibited the poorest performance, with RMSE = 5.46 m and Mean_E = 4.42 m. Wild-GS failed to generate valid outputs.

Dortmund It highlights reconstruction performance on dense urban blocks with high structural complexity. Photogrammetry achieved the highest accuracy, with RMSE = 0.88 m, MAE = 0.68 m, STD = 0.55 m, and Mean_E = 0.75 m, demonstrating superior geometric precision and stability across diverse structural patterns. Among 3DGS approaches, Octree-GS again achieved the best results, with RMSE = 1.63 m and MAE = 1.07 m, confirming the effectiveness of its hierarchical representation. CityGaussian reached a competitive RMSE of 1.77 m but showed instability reflected by higher variance. GOF and BlockGaussian achieved moderate accuracy but presented larger errors in regions with intricate geometry, while Wild-GS produced the poorest performance, with RMSE exceeding 6.8 m and Mean_E = 5.75 m. These findings further emphasize traditional photogrammey has better accuracy in urban geometric reconstruction than 3DGS methods.

MatrixCity - Block 2 It focuses on block-level reconstruction within a synthetic city. Octree-GS once more achieved the best results across all metrics with RMSE equal to 0.73m, MAE equal to 0.53m, STD equal to 0.49m, and Mean_E equal to 0.61m, which confirms its robustness across different urban scales. BlockGaussian and CityGaussian provided reasonable reconstructions with RMSE between 1.7m and 1.8m, but with higher noise compared to Octree-GS. GOF performed poorly with RMSE equal to 3.43m and STD equal to 2.9m, and Wild-GS failed to converge.

5.1.3 Geometric Evaluation of fine-structures

To evaluate the fine-structure reconstruction capability of Gaussian-based methods compared with traditional photogrammetry, we selected several detailed regions from the Dortmund dataset. As shown in Table 4, CityGaussian achieved the best overall performance, with an RMSE of 0.43m, MAE of 0.34m, and STD of 0.26m, demonstrating its strong ability to preserve fine geometric details. Octree-GS followed closely, achieving an RMSE of 0.48, MAE of 0.41m, and STD of 0.26m, indicating consistent accuracy and stability. GOF produced moderate results, as its dense outputs introduced excessive attached noise points. In contrast, Photogrammetry failed to reconstruct those small structures due to the limited GSD of the images.

5.2 Graphic Evaluation

5.2.1 Qualitative and Quantitative Evaluation of Rendered Images

We qualitatively and quantitatively evaluate the rendering quality of the 3DGS methods across three datasets. As shown in Figure 2 and Table 5, BlockGaussian consistently achieves superior performance in photometric metrics, obtaining the highest SSIM and PSNR values while maintaining the lowest LPIPS, indicating sharp structural preservation and faithful texture recovery. Octree-GS and CityGaussian perform comparably overall, but their relative strengths vary across datasets. On the UseGeo dataset, Octree-GS slightly outperforms CityGaussian while on MatrixCity Block_2, CityGaussian shows superior metrics. In contrast, GOF exhibits unstable performance across datasets. It tends to generate noisy points near scene boundaries, which can cause severe distortions in certain rendering views, leading to lower photometric scores compared to BlockGaussian and the other two methods. Wild-GS exhibits the weakest performance. Its rendered images show severe RGB mismatches with the GT, along with multiple local distortions. Fine texture and structural details are obviously lost, resulting in significantly lower SSIM and PSNR, as well as higher LPIPS, which further

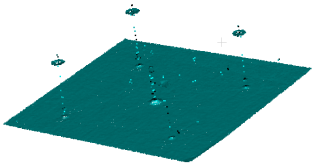
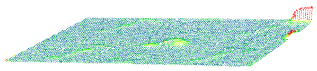
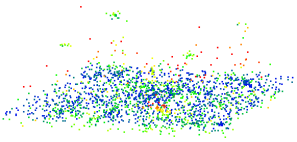
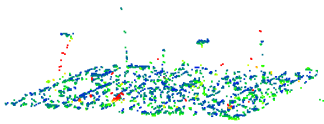
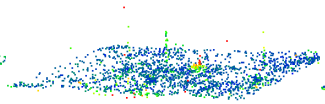
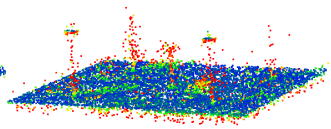
GT		Photogrammetry	BlockGaussian
			
RMSE MAE STD		Failed	1.59 1.17 1.07
Octree-GS	**CityGaussian**	**GOF**	**Wild-GS**
			Failed
0.48 0.41 0.26	0.43 0.34 0.26	0.70 0.47 0.52	Failed

Table 4. Geometric evaluation [m] of fine-structure reconstruction between 3DGS methods and traditional photogrammetry.

confirms its limited generalization and unsuitability for urban scene reconstruction.

Dataset	Method	SSIM ↑	PSNR ↑	LPIPS ↓
UseGeo	BlockGaussian	**0.8229**	**28.28**	**0.242**
	Octree-GS	0.6820	24.45	0.265
	CityGaussian	0.5154	21.01	0.853
	Gaussian Opacity Fields	0.3693	17.20	0.627
	Wild-GS	–	–	–
Dortmund	BlockGaussian	**0.9626**	**32.90**	**0.049**
	Octree-GS	0.8299	26.56	0.150
	CityGaussian	0.8554	28.10	0.144
	Gaussian Opacity Fields	0.9092	30.29	0.094
	Wild-GS	0.7072	25.08	0.335
MatrixCity	BlockGaussian	**0.9647**	**37.46**	**0.146**
	Octree-GS	0.7374	21.49	0.338
	CityGaussian	0.9007	32.64	0.214
	Gaussian Opacity Fields	0.8134	30.07	0.390
	Wild-GS	–	–	–

Table 5. Photometric evaluation metrics on three tested datasets. Higher SSIM/PSNR and lower LPIPS are better.

5.2.2 Quantitative Evaluation of GPU Memory Usage

Table 6 summarizes the peak GPU memory usage of all methods during training. On Dortmund, GOF and Wild-GS consume more memory, while BlockGaussian, Octree-GS, and CityGaussian remain more efficient. For UseGeo and MatrixCity, the memory demand increases for all methods due to more training images, with GOF and Wild-GS again showing the highest consumption. Octree-GS achieves notable efficiency through its hierarchical octree structure, which minimizes redundancy. BlockGaussian further reduces memory and computation by partitioning scenes into smaller blocks. CityGaussian saves memory via aggressive Gaussian pruning but sacrifices reconstruction completeness. In contrast, Wild-GS requires substantial memory on large datasets, limiting its scalability. Overall, these results reveal a trade-off between computational efficiency and reconstruction quality. Octree-GS and CityGaussian perform best in global and fine-structure reconstruction, respectively, while

Method	UseGeo	Dortmund	MatrixCity
BlockGaussian	24108	5626	14520
Octree-GS	**18964**	6043	11403
CityGaussian	23814	**4025**	**6212**
GOF	42096	17727	39746
Wild-GS	32845	8027	25195

Table 6. GPU memory usage (MB) for each method during training on the Dortmund and MatrixCity dataset. Since GPU usage is dynamic, we pick the peak values as the result.

BlockGaussian excels in rendering metrics. Considering both geometry and photorealism, Octree-GS emerges as the most balanced method, demonstrating the strength of hierarchical splatting for urban-scale 3D reconstruction.

6 Conclusions

This work presented a systematic evaluation of recent 3DGS methods for large-scale urban scene reconstruction, comparing them with traditional photogrammetry and LiDAR data as ground truth. Our study examines 3DGS approaches in terms of scalability, generalization, and efficiency in complex urban environments, where reconstruction accuracy, detail fidelity, and memory usage are critical. Experiments on five 3DGS methods show that BlockGaussian, CityGaussian, GOF, and Octree-GS successfully reconstruct all evaluated datasets, whereas Wild-GS produces valid output only on the Dortmund dataset. In general, no 3DGS method dominates in all aspects of reconstruction and rendering. Traditional photogrammetry achieves the highest accuracy in global geometric reconstruction, but struggles with fine or reflective structures. CityGaussian performs best in fine-structure reconstruction, BlockGaussian achieves best results in image rendering, and Octree-GS emerges as the most balanced method, achieving strong geometric accuracy alongside competitive rendering quality. These results provide valuable insights for researchers by offering guidance on method selection, computational resource allocation, and performance trade-offs across

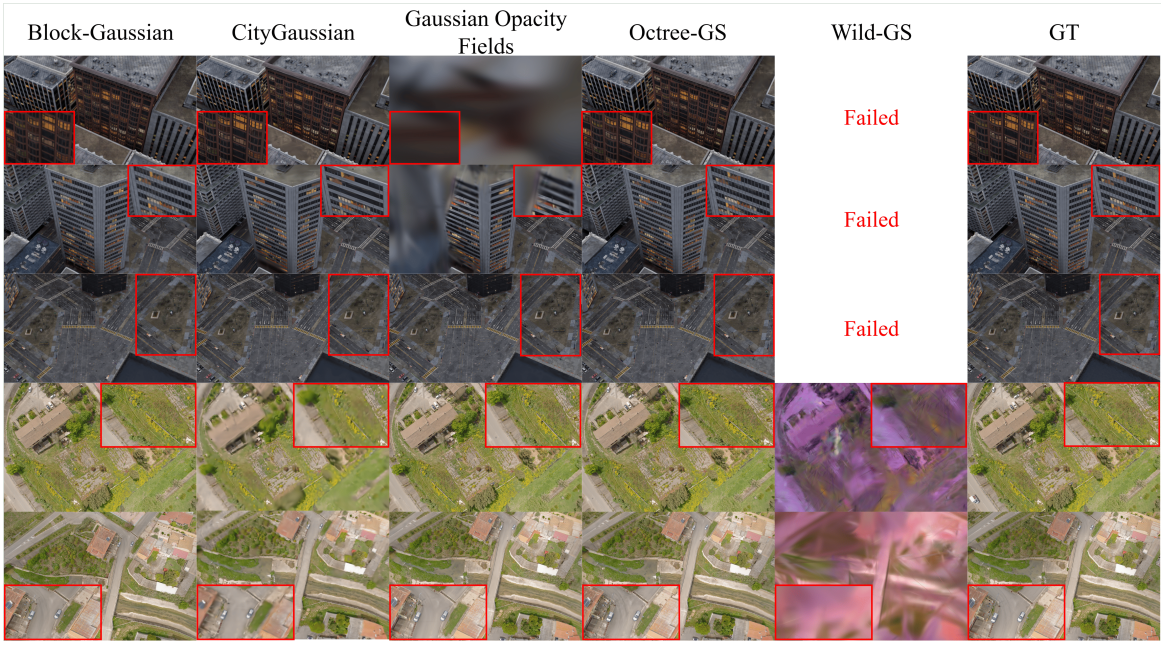


Figure 2. Qualitative graphic-based comparison of the evaluated 3DGS methods across MatrixCity (top) and UseGeo (bottom) datasets. Fine structures such as wires and windows are highlighted with close-ups for clearer visibility.

diverse application scenarios. We believe that these findings will facilitate future research on urban-scale 3D reconstruction leveraging 3DGS techniques, ideally integrated with conventional photogrammetric MVS approaches.

References

Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S. M., Szeliski, R., 2011. Building rome in a day. *Communications of the ACM*, 54(10), 105–112.

Akhavi Zadegan, A., Vivet, D., Hadachi, A., 2025. Challenges and advancements in image-based 3D reconstruction of large-scale urban environments: a review of deep learning and classical methods. *Frontiers in Computer Science*, 7, 1467103.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., 2023. Zip-nerf: Anti-aliased grid-based neural radiance fields. *ICCV*, 19697–19705.

Chen, J., Ye, W., Wang, Y., Chen, D., Huang, D., Ouyang, W., Zhang, G., Qiao, Y., He, T., 2024. Gigags: Scaling up planar-based 3d gaussians for large scene surface reconstruction. *arXiv preprint arXiv:2409.06685*.

Dong, Z., Chen, K., Lv, Z., Yu, H.-X., Zhang, Y., Zhang, C., Zhu, Y., Tian, S., Li, Z., Moffatt, G. et al., 2025. Digital twin catalog: A large-scale photorealistic 3d object digital twin dataset. *CVPR*, 753–763.

Feng, G., Chen, S., Fu, R., Liao, Z., Wang, Y., Liu, T., Hu, B., Xu, L., Pei, Z., Li, H. et al., 2025. Flashgs: Efficient 3d gaussian splatting for large-scale and high-resolution rendering. *CVPR*, 26652–26662.

Furukawa, Y., Ponce, J., 2009. Accurate, dense, and robust multiview stereopsis. *IEEE TPAMI*, 32(8), 1362–1376.

Gao, Y., Li, H., Chen, J., Zou, Z., Zhong, Z., Zhang, D., Sun, X., Han, J., 2025. Citygs-x: A scalable architecture for efficient and geometrically accurate large-scale scene reconstruction. *arXiv preprint arXiv:2503.23044*.

Guédon, A., Lepetit, V., 2024. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. *CVPR*, 5354–5363.

Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S., 2024. 2d gaussian splatting for geometrically accurate radiance fields. *ACM SIGGRAPH*, 1–11.

Huang, J., Stoter, J., Peters, R., Nan, L., 2022. City3D: Large-scale building reconstruction from airborne LiDAR point clouds. *Remote Sensing*, 14(9), 2254.

Jiang, L., Ren, K., Yu, M., Xu, L., Dong, J., Lu, T., Zhao, F., Lin, D., Dai, B., 2025. Horizon-gs: Unified 3d gaussian splatting for large-scale aerial-to-ground scenes. *CVPR*, 26789–26799.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4), 139–1.

Li, Y., Jiang, L., Xu, L., Xiangli, Y., Wang, Z., Lin, D., Dai, B., 2023. Matrixcity: A large-scale city dataset for city-scale neural rendering and beyond. *ICCV*, 3205–3215.

Li, Z., Zhang, Y., Wu, C., Zhu, J., Zhang, L., 2024. Ho-gaussian: Hybrid optimization of 3d gaussian splatting for urban scenes. *ECCV*, Springer, 19–36.

Liao, M., Dong, H. B., Wang, X., Yan, Z., Shao, Y., 2025. GM-MoE: Low-Light Enhancement with Gated-Mechanism Mixture-of-Experts. *arXiv preprint arXiv:2503.07417*.

Liao, Y., Xie, J., Geiger, A., 2022. Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE TPAMI*, 45(3), 3292–3310.

Lin, J., Li, Z., Tang, X., Liu, J., Liu, S., Liu, J., Lu, Y., Wu, X., Xu, S., Yan, Y. et al., 2024. Vastgaussian: Vast 3d gaussians for large scene reconstruction. *CVPR*, 5166–5175.

Lin, L., Liu, Y., Hu, Y., Yan, X., Xie, K., Huang, H., 2022. Capturing, reconstructing, and simulating: the urbanscene3d dataset. *ECCV*, Springer, 93–109.

Liu, Y., Luo, C., Fan, L., Wang, N., Peng, J., Zhang, Z., 2024a. Citygaussian: Real-time high-quality large-scale scene rendering with gaussians. *ECCV*, Springer, 265–282.

Liu, Y., Luo, C., Mao, Z., Peng, J., Zhang, Z., 2024b. Citygaussianv2: Efficient and geometrically accurate reconstruction for large-scale scenes. *arXiv preprint arXiv:2411.00771*.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., Ng, R., 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106.

Rematas, K., Liu, A., Srinivasan, P. P., Barron, J. T., Tagliasacchi, A., Funkhouser, T., Ferrari, V., 2022. Urban radiance fields. *CVPR*, 12932–12942.

Remondino, F., Karami, A., Yan, Z., Mazzacca, G., Rigon, S., Qin, R., 2023. A critical analysis of NeRF-based 3D reconstruction. *Remote Sensing*, 15(14), 3585.

Ren, K., Jiang, L., Lu, T., Yu, M., Xu, L., Ni, Z., Dai, B., 2024. Octree-gs: Towards consistent real-time rendering with lod-structured 3d gaussians. *arXiv preprint arXiv:2403.17898*.

Roman, O., Bassier, M., Ricciuti, S., Farella, E. M., Remondino, F., Viesi, D., 2025. Digital Twins and CFD simulations for accurate sensor positioning. *ISPRS Archives*, 48, 1291–1298.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. *CVPR*, 1, IEEE, 519–528.

Shao, Y., He, H., Li, S., Chen, S., Long, X., Zeng, F., Fan, Y., Zhang, M., Yan, Z., Ma, A. et al., 2025a. EventVAD: Training-Free Event-Aware Video Anomaly Detection. *arXiv preprint arXiv:2504.13092*.

Shao, Y., Liang, S., Ling, Z., Yan, M., Liu, H., Chen, S., Yan, Z., Zhang, C., Qin, H., Magno, M. et al., 2024a. GWQ: Gradient-Aware Weight Quantization for Large Language Models. *arXiv preprint arXiv:2411.00850*.

Shao, Y., Lin, D., Zeng, F., Yan, M., Zhang, M., Chen, S., Fan, Y., Yan, Z., Wang, H., Guo, J. et al., 2025b. TR-DQ: Time-Rotation Diffusion Quantization. *arXiv preprint arXiv:2503.06564*.

Shao, Y., Lin, X., Long, X., Chen, S., Yan, M., Liu, Y., Yan, Z., Ma, A., Tang, H., Guo, J., 2025c. ICM-Fusion: In-Context Meta-Optimized LoRA Fusion for Multi-Task Adaptation. *arXiv preprint arXiv:2508.04153*.

Shao, Y., Xu, Y., Long, X., Chen, S., Yan, Z., Yang, Y., Liu, H., Wang, Y., Tang, H., Lei, Z., 2024b. AccidentBlip: Agent of Accident Warning based on MA-former. *arXiv preprint arXiv:2404.12149*.

Shao, Y., Yan, M., Liu, Y., Chen, S., Chen, W., Long, X., Yan, Z., Li, L., Zhang, C., Sebe, N. et al., 2025d. In-Context Meta LoRA Generation. *arXiv preprint arXiv:2501.17635*.

Snavely, N., Seitz, S. M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3d. *ACM siggraph*, 835–846.

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P. P., Barron, J. T., Kretzschmar, H., 2022. Block-nerf: Scalable large scene neural view synthesis. *CVPR*, 8248–8258.

Turki, H., Ramanan, D., Satyanarayanan, M., 2022. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. *CVPR*, 12922–12931.

Wang, F., Galliani, S., Vogel, C., Speciale, P., Pollefeys, M., 2021. Patchmatchnet: Learned multi-view patchmatch stereo. *CVPR*, 14194–14203.

Wang, N., Chen, Y., Xiao, L., Xiao, W., Li, B., Chen, Z., Ye, C., Xu, S., Zhang, S., Yan, Z. et al., 2025a. Unifying Appearance Codes and Bilateral Grids for Driving Scene Gaussian Splatting. *arXiv preprint arXiv:2506.05280*.

Wang, T., Wang, X., Hou, Y., Xu, Y., Zhang, W., Zhan, Z., 2025b. PG-SAG: Parallel Gaussian Splatting for Fine-Grained Large-Scale Urban Buildings Reconstruction via Semantic-Aware Grouping. *PFG*, 1–16.

Wu, Y., Qi, Z., Shi, Z., Zou, Z., 2025. BlockGaussian: Efficient Large-Scale Scene Novel View Synthesis via Adaptive Block-Based Gaussian Splatting. *arXiv preprint arXiv:2504.09048*.

Xu, J., Mei, Y., Patel, V., 2024. Wild-gs: Real-time novel view synthesis from unconstrained photo collections. *NIPS*, 37, 103334–103355.

Yan, Z., Dong, W., Shao, Y., Lu, Y., Liu, H., Liu, J., Wang, H., Wang, Z., Wang, Y., Remondino, F. et al., 2025a. Renderworld: World model with self-supervised 3d label. *ICRA*, 6063–6070.

Yan, Z., Li, L., Shao, Y., Chen, S., Wu, Z., Hwang, J.-N., Zhao, H., Remondino, F., 2024. 3dsceneeditor: Controllable 3d scene editing with gaussian splatting. *arXiv preprint arXiv:2412.01583*.

Yan, Z., Mazzacca, G., Rigon, S., Farella, E. M., Trybala, P., Remondino, F. et al., 2023. NeRFBK: a holistic dataset for benchmarking NeRF-based 3D reconstruction. *ISPRS Archives*, 48(1), 219–226.

Yan, Z., Padkan, N., Trybała, P., Farella, E. M., Remondino, F., 2025b. Learning-Based 3D Reconstruction Methods for Non-Collaborative Surfaces—A Metrological Evaluation. *Metrology*, 5(2), 20.

Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth Anything V2. *arXiv:2406.09414*.

Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L., 2018. Mvsnet: Depth inference for unstructured multi-view stereo. *ECCV*, 767–783.

Yu, Z., Gao, S., 2020. Fast-mvsnet: Sparse-to-dense multi-view stereo with learned propagation and gauss-newton refinement. *CVPR*, 1949–1958.

Yu, Z., Sattler, T., Geiger, A., 2024. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM TOG*, 43(6), 1–13.

Zhang, H., Yao, Y., Xie, K., Fu, C.-W., Zhang, H., Huang, H., 2021. Continuous aerial path planning for 3D urban scene reconstruction. *ACM TOG*, 40(6), 225–1.

Zhou, H., Shao, J., Xu, L., Bai, D., Qiu, W., Liu, B., Wang, Y., Geiger, A., Liao, Y., 2024. Hugs: Holistic urban 3d scene understanding via gaussian splatting. *CVPR*, 21336–21345.