Nation Scale NeRF Reconstruction

Alexander Mai[1], Scott McAvoy[1,3], Jonathan Klingspon[1], Davide Forcellini[2], Falko Kuester[1]

[1] Cultural Heritage Engineering Initiative, University of California San Diego – (atm008, smcavoy, jklingspo, fkuester)@ucsd.edu
[2] University of the Republic of San Marino – davide.forcellini@unirsm.sm
[3] OpenHeritage3D.org – (smcavoy@oh3d.org)

Keywords: NeRF, AI Reconstruction, Multi-Cam Photogrammetry, Data-fusion

Abstract

Neural Radiance Field (NeRF) rendering methodologies provide 3D reconstructions which better represent features and characteristics common to built heritage, which are otherwise poorly represented by traditional structure from motion reconstruction and rendering techniques. It is currently limited in its ability to scale to large projects. This paper, through a case study involving large scale seismic vulnerability surveys in San Marino, investigates and proposes systems which can extend NERF rendering to large multiscalar datasets linked by geographic coordinates.

1. Introduction

The creation of photorealistic, navigable digital replicas of the real world is a long-standing goal in computer graphics and vision. Such digital twins have profound applications, ranging from cultural heritage preservation and urban planning to providing training grounds for autonomous systems and creating immersive content for virtual and augmented reality. Achieving this level of realism has historically been challenging, often requiring complex manual modeling to account for the shortcomings of traditional photogrammetric reconstruction. The introduction of Neural Radiance Fields (NeRFs) marked a paradigm shift in this domain, offering a new method for synthesizing photorealistic novel views from a collection of input images.

A NeRF represents a continuous, three-dimensional scene as a lightweight neural network. This network acts as a function that maps a 3D spatial coordinate (x, y, z) and a 2D viewing direction (θ, ϕ) to the color and volume density at that point. By querying this network along camera rays and using classical volume rendering techniques, NeRF can generate highly realistic images from previously unseen viewpoints. This approach excels at capturing complex geometric details and view-dependent effects like reflections and translucency with unprecedented fidelity. Practically speaking, within in the built environment, features like windows, wires, pipes, flat mono-color walls, and thin metal structures are better represented through NeRF.

Despite their quality, the original NeRF models were hindered by slow training and rendering times, limiting their practical use. This spurred a wave of research aimed at accelerating the paradigm, leading to a new class of methods that combine neural representations with explicit geometric primitives. These techniques, which can render scenes in real-time, have become the new state-of-the-art. This paper rigorously evaluates these advanced methodologies through a comprehensive digital reconstruction of the entire nation-state of San Marino. Unlike previous studies that often focused on smaller, object-centric scenes, our evaluation leverages a massive, high-quality dataset to test the limits of these methods on a national scale (Martin-Brualla et al., 2021). Specifically, we investigate cutting-edge techniques including 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023), its Markov Chain Monte Carlo variant (3DGS-MCMC) (Kheradmand et al., 2024), EVER (Mai et al., 2024), Deformable Beta Splatting (DBS) (Liu et al., 2025), and SVRaster (Sun et al., 2024). The goal is to identify the scalability, robustness, and failure points of these reconstruction methods, thereby contributing valuable insights into the future of large-scale digital twinning and heritage preservation.

The Republic of San Marino is a 61 square kilometer nation-state, and the fifth smallest country in the world. Though relatively small, it represents a sufficient scale, density, and variety of topographies and urban plans upon which we can plan systems of unlimited scalability. San Marino's entire city center and market town of Borgo are designated as UNESCO world heritage sites, and are "at risk" due to seismological threats (Forcellini 2016). This nation presents unique challenges and opportunities for 3D documentation and visualization due to its densely packed vertical architecture and complex topography. Medieval buildings are stacked up on a steep slope, with some of the most important heritage sites perched precariously on a sheer cliffside (figures 1 and 2). Tight streets, in the city center, create challenging subjects for mapping, as the upper floors of buildings must be viewed from steep upwards angles (figure 3). Though currently limited in metrological applications, Nerf rendering offers an attractive means to visualize heritage structures. Both built and natural heritage contains features which do not reconstruct accurately within traditional mesh, point, and depth-map based structure from motion (SFM) pipelines. Reflective features like windows and water surfaces, featureless surfaces like clean plaster walls, and thin features like iron rods, power lines, vegetation etc. . . are often represented as globular deformed masses.

2. Relevant Work

The digital documentation of cultural heritage has traditionally relied on well-established photogrammetry pipelines, such as Structure-from-Motion (SfM) combined with Multi-View Stereo (MVS). These methods are valued for their high geometric accuracy (Schonberger and Frahm, 2016). However, they often struggle to produce faithful reconstructions of sites with challenging materials, such as reflective, translucent, or texture-less surfaces, which are common in heritage artifacts and architecture (Remondino et al., 2023).

The introduction of Neural Radiance Fields (NeRF) marked a significant shift, offering a new paradigm for creating photorealistic visualizations from image collections. A growing body of research has conducted direct comparisons between NeRF and traditional photogrammetry for heritage documentation (Clini et al., 2024). Previous studies have found that NeRF-based methods excel at capturing the visual appearance of challenging materials and can often produce more complete and visually coherent models, especially when dealing with a limited number of input images or lower-resolution data. However, this visual fidelity often comes at the cost of geometric precision, with NeRF-derived models typically exhibiting more noise and lower accuracy compared to the outputs of mature photogrammetric pipelines (Clini et al., 2024).

As research progressed beyond single objects to large-scale environments, methods were developed to handle expansive scenes, such as dividing a large area into manageable sections for reconstruction (Tancik et al., 2022). However, the most significant recent development for the heritage field has been the introduction of 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023). 3DGS was the first rasterization based NeRF to achieve high quality. Rather than iterating over pixels, tracing through the 3D scene, 3DGS iterates over the 3D splats, projecting them onto the screen. Several recent comparative studies focusing specifically on heritage sites have evaluated NeRF against 3DGS. A clear consensus from this research is that 3DGS offers a dramatic advantage in efficiency, with significantly faster training and real-time rendering speeds (Atik, 2025). Furthermore, studies show that 3DGS often produces reconstructions with less noise, sharper details, and superior texture quality compared to NeRF (Clini et al., 2024).

While both NeRF and 3DGS represent a major leap forward for the visualization of cultural heritage, a critical conclusion from the comparative literature is that neither technology has yet achieved the level of geometric accuracy required for metrological documentation (Clini et al., 2024). For applications demanding precise measurements, traditional photogrammetry remains the more reliable method. Therefore, current research suggests that neural rendering techniques like NeRF and 3DGS are best seen as powerful complementary tools for visualization, virtual tourism, and capturing the aesthetic qualities of a site, rather than as a complete replacement for established high-accuracy survey methods (Mazzacca et al., 2023).

3. The Data

Our datasets, collected by expert operators for the purpose of multi-scale seismological studies, includes objects sourced from a comprehensive 9 year survey project (Lo et al., 2023). The multi-modal dataset incorporates satellite stereo imagery, drone surveys, over 1000 terrestrial LiDAR scans with corresponding high dynamic range (HDR) image spheres, and hundreds of thousands of images from 33mp cameras and a specialized multicam (Meyer et al., 2020). It comprises a 360-degree panoramic dataset capturing the distinctive topographical layout of San Marino, situated atop Mount Titano, complemented by extensive detailed street-level and interior building data for key monumental structures. This controlled, high-quality dataset enables comprehensive testing and detailed characterization of reconstruction performance under realistic yet optimal conditions, and enables a comparative analysis across scales, data modalities, and various other conditions impacting reconstruction and rendering from optical data sources.

The majority of our data is made available to the public through an open-data repository specifically designed for the re-use of 3d data (McAvoy et al., 2023), including both traditional outputs suitable for metrological study, as well as pre-packaged images and alignment derivatives enabling Nerf reconstruction through external rendering pipelines (McAvoy et al., 2024). In this way, we empower other scholars to reproduce and extend our results.

Street level imagery is provided through the Looq, a multi-cam photogrammetry mobile mapping system. The dataset includes 27,000 geolocated images, or 6750 image multicam image sets, captured in 9 surveys performed at various times of day over a 4 day period (McAvoy et al., 2025).

4. Method

4.1 Method Comparison

Table 1 provides a comparative analysis of the various state-of-the-art real-time radiance field techniques. These methods are differentiated by their choice of geometric primitive, appearance representation, and core training heuristics.

A primary distinction lies in the densification strategy used during training. 3D Gaussian Splatting (3DGS), EVER, and SVRaster employ a gradient-based approach, accumulating loss to guide the addition of new primitives in regions of high error. In contrast, 3DGS-MCMC and Deformable Beta-Splatting (DBS) utilize a probabilistic MCMC-based method to propose the addition or removal of primitives. This treats densification as a sampling problem rather than a direct response to error, leading to a different evolution of the scene representation.

The rendering pipelines are also a key factor. 3DGS, 3DGS-MCMC, DBS, and SVRaster are all fundamentally rasterization-based, using a forward splatting process to project primitives onto the screen. In fact, 3DGS and 3DGS-MCMC share the exact same rendering code. DBS and SVRaster use similar pipelines, adapted for their specific primitives (beta kernels and voxels, respectively). This shared approach contrasts sharply with EVER, which is a ray-traced method that casts rays to sample its ellipsoidal primitives, a more computationally intensive process.

Regarding appearance, most methods use standard Spherical Harmonics (SH) to model view-dependent effects. DBS is an exception, employing a custom spherical representation that can capture higher-frequency lighting details than typical low-order SH, though it operates under an assumption of sparsity.

These design choices directly impact performance. The rasterization based methods are significantly faster. DBS leads with the highest rendering speed (138 FPS) and fastest training (10 minutes), owing to its efficient beta kernels and MCMC strategy. Since 3DGS and 3DGS-MCMC use identical renderers, the performance difference (131 vs 93 FPS) can be attributed entirely to the scene representation produced by their respective densification strategies; the MCMC-generated scenes are inherently more complex to render. SVRaster's high performance (121 FPS) is also explained by its efficient voxel rasterization pipeline. Finally, EVER's reliance on ray tracing explains its comparatively lower framerate (20 FPS) and the longest training time (90 minutes).

We finally included one additional method: Hierarchical 3DGS. This method builds on 3DGS, but adds correct anti-aliasing of

Feature	3DGS	3DGS-MCMC	EVER	DBS	SVRaster	Hier 3DGS
Loss Accumulation	✓	×	✓	×	✓	✓
MCMC	×	✓	×	✓	×	×
Sharp Primitives	×	×	✓	✓	✓	×
Spherical Harmonics	✓	✓	✓	×	✓	✓
Volume Rendering	×	×	✓	×	✓	×
Primitive	Gaussian	Gaussian	Ellipsoid	Beta Kernel	Voxel	Gaussian
Training Speed	16m	16m	90m	10m	30m	336m
FPS	131	93	20	138	121	≈ 30
Depth Regularizer	×	×	×	×	×	✓
Chunking	×	×	×	×	×	✓
LODs	×	×	×	×	×	✓

Table 1. FPS taken from (Sun et al., 2024) and (Liu et al., 2025). Rescaled to represent performance on a NVIDIA RTX3090. Training numbers represent speed on an NVIDIA RTX4090.

primitives, depth supervision using Depth Anywhere V2 (Yang et al., 2024), levels of detail (LODs), and hierarchical reconstruction, which are all aimed at improving it's handling of large datasets. The combination of chunking and LODs enable it to handle far more primitives, while also improving densification. Depth supervision helps on more challenging datasets. For the San Marino streets dataset, pre-processing steps took 2 hours, with an additional 5.6 hours of training.

4.2 Tuning

The adaptation of radiance field methods to large, complex datasets necessitates tuning of the training and densification schedule. For Gaussian-based methods like 3DGS and 3DGS-MCMC, a more conservative densification strategy is crucial. On large scenes, aggressive and early densification can lead to a premature and excessive proliferation and pruning of Gaussians, especially if pruning takes place before every image has been seen. By delaying the start of densification (e.g., `--densify_from_iter 8000`), increasing the interval between densification steps (`--densification_interval 1000`), and concluding the process well before training ends (`--densify_until_iter 30000`), the model is encouraged to first optimize the positions and properties of existing primitives. This slower schedule, combined with a lower spherical harmonic degree (`--sh_degree 1`), ensures more stable training and better resource management over a higher total number of iterations (`--iterations 60000`).

For voxel-based methods such as SVRaster, we accomplish similar objectives with the following parameters: `sche_mult: 3`, `adapt_from: 3000`, `prune_from: 3000`). Since it also offers options to delay regularization, we delay those to prevent them from penalizing the model too harshly during the initial learning phase (`dist_from: 3000`, `tv_from: 3000`).

We did not tune Hierarchical 3DGS, as it was designed to handle large scenes.

4.3 Metrics

Our two datasets do not have a consistent exposure and white balance, which is known to make evaluation difficult. Part of the problem is that the multicam strips the metadata from the images, leaving no information about what these exposure and white balance values were.

GLO vectors (Duckworth et al., 2024) and Bilateral Grids (Wang et al., 2024b) are typical solutions to this problem. However, GLO vectors are not agnostic to the rendering method and the application of GLO vectors to a primitive based rendering scheme is hacky. We tried Bilateral Grids, but they do not scale well in the number of images in the dataset, bringing training times from minutes to days.

As a result, we used no mitigation strategy, in the same vein as the standard evaluation of Tanks and Temples (Knapitsch et al., 2017). The lack of metadata meant that any mitigation strategy would prevent evaluation. Instead, we relied on the invariance of the test metrics to white balance and exposure, as both LPIPs (Zhang et al., 2018) and SSIM (Wang et al., 2004) have some resilience against these problems.

5. Results

The images were split into train and test portions by selecting every 8th image for the test set. The results in the Table 2 represent the metrics measured on the test set, after training the model on the training set. It is important to remember that Hierarchical 3DGS is in a whole other category of method, as it uses extremely powerful depth regularizers and chunking to help handle the larger scene.

Let's first talk about the qualitative results for the San Marino streets dataset. The only method that was able to reconstruct a result that resembles the ground truth was EVER. 3DGS, 3DGS-MCMC, Hier-3DGS, and SVRaster all produce results that capture a few details of the original dataset. DBS, however, completely failed to capture any aspects of the original dataset. This shows that the MCMC densification strategy, although highly successful for the Mip-NeRF360 datasets, does not seem to be robust. Interestingly, Hierarchical 3DGS actually appears to be less robust than the original 3DGS. As seen in Fig. 2, when Hierarchical 3DGS manages to converge on an area, it has the potential to get the best results. However, it also failed to reconstruct many chunks.

Quantitative results show how untrustworthy PSNR can be. To the human eye, the fact that the structure of the images reconstructed by EVER resembles the ground truth outweighs the inaccuracies in the coloration. However, to PSNR, the reverse is true. This is not really a surprising result, as this is the main

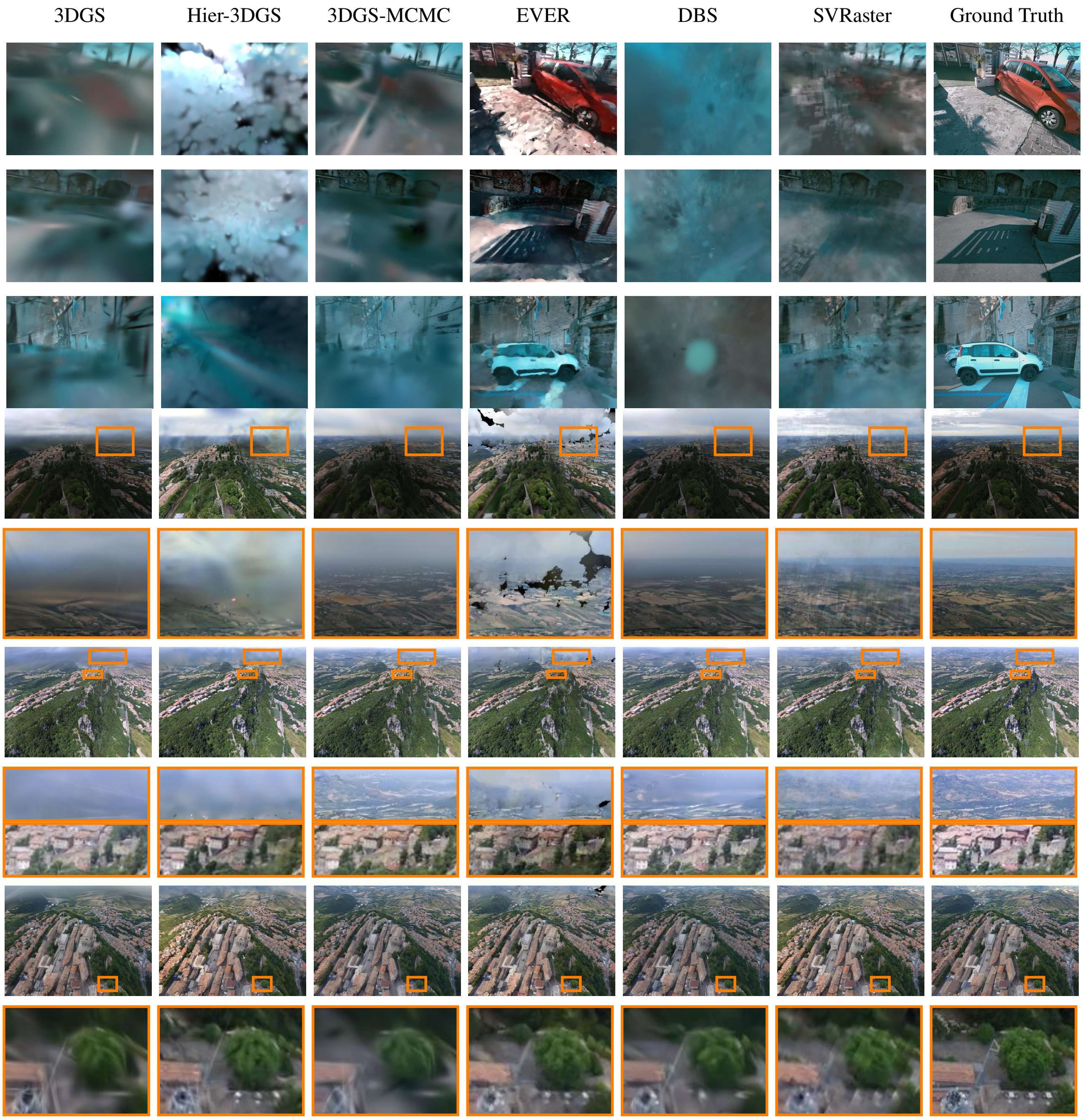

Figure 1. Qualitative comparison on San Marino datasets. Images from test set shown. Top three rows: street level reconstruction of San Marino using multicam setup (Meyer et al., 2020). Bottom three rows: aerial reconstruction using DJI drone.

reason datasets with varying exposure make for difficult evaluation. However, LPIPs and SSIM both retrieve an ordering more in line with the perception of the human eye. This indicates that it might be possible to evaluate methods on imperfect datasets, as long as LPIPs and SSIM are the focus, instead of PSNR.

The San Marino aerial dataset reconstruction was more successful, with all methods retrieving a decent result. What is notable is the black artifacts in EVER, caused by early termination of the rays, and the blurriness of the distant objects from 3DGS. The results on the San Marino aerial dataset demonstrate the results from 3DGS with correct antialiasing, which was not in the 3DGS or 3DGS-MCMC results. As can be seen in the qualitative evaluation, this results in improved detail on the roof of some buildings on the 6th row. EVER, however, seem to achieve the sharpest results. Even with the black artifacts, it achieves very high rankings on the perceptual metrics.

These datasets show a stark contrast to the standard Mip-NeRF360 (Barron et al., 2022) evaluation. The Mip-NeRF360 dataset was originally designed to demonstrate the test the capabilities of early view synthesis work. Now that methods have grown more capable, it seems likely that this dataset is no longer sufficient for testing how well these NeRFs will work in the wild. The Mip-NeRF 360 dataset does not have a large variety of different capture approaches. All were taken with a consumer camera from multiple angles. The San Marino Aerial capture is most similar to this, which explains why it worked reasonably well. However, the San Marino Streets capture, being taken from a multicam carried at the same height, is starkly different. This

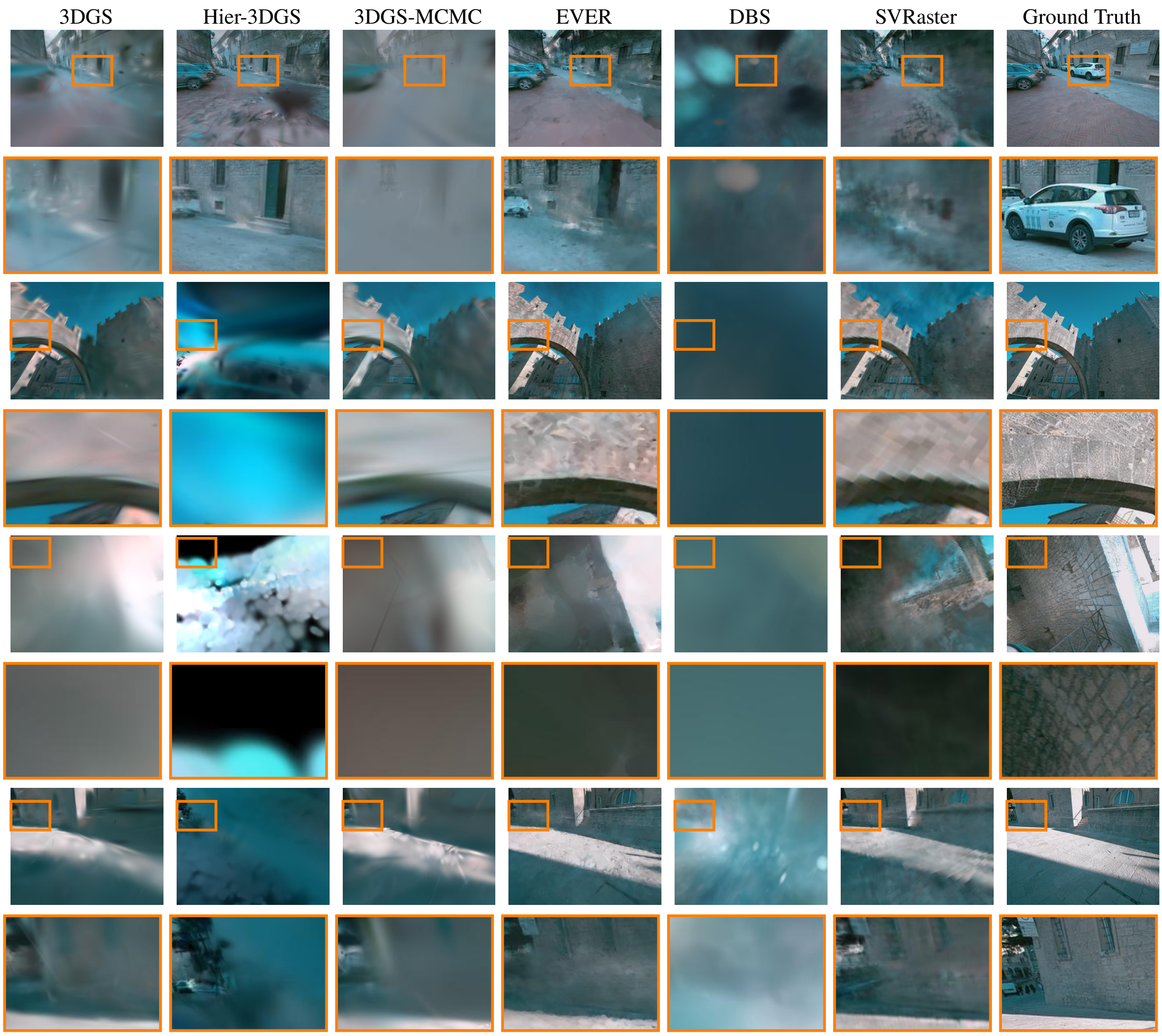

Figure 2. Qualitative results on the train set of images to help illustrate the main challenges with large scale captures. Row one: Moving object, like the car. Row two: resolution problems, often solved with LODs or chunking. Row three: mediocre capture. Poor capture of the floor leaks upwards, degrading the building reconstruction. Row four: exposure change. Dark part of image that auto exposure targets highlighted.

	San Marino Street					
	3DGS	3DGS-MCMC	EVER	DBS	SVRaster	Hier-3DGS
PSNR ↑	19.57	18.73	18.42	14.71	18.17	14.9
SSIM ↑	.501	.491	.522	.453	.474	0.4325
LPIPS C ↓	.710	.716	.633	.790	.700	0.67696
LPIPS ↓	.666	.666	.605	.725	.676	0.63449
	San Marino Aerial					
	3DGS	3DGS-MCMC	EVER	DBS	SVRaster	Hier-3DGS
PSNR ↑	22.78	22.56	19.19	24.40	22.87	21.61804
SSIM ↑	.702	.724	.730	.774	.742	0.7096
LPIPS C ↓	.402	.377	.280	.310	.282	0.33927
LPIPS ↓	.363	.335	.259	.273	.259	0.30969
	Mip-NeRF 360					
	3DGS	3DGS-MCMC	EVER	DBS	SVRaster	Hier-3DGS
PSNR ↑	27.45	28.08	27.51	28.75†	27.33†	-
SSIM ↑	.815	.837	.825	0.845†	0.822†	-
LPIPS C ↓	.257	.177	.233		0.219†	-
LPIPS ↓	.216	.179	.194	0.179†		-

Table 2. Quantitative result. Results with † were taken from their original papers. Others were recalculated. LPIPs (C) is the corrected LPIPs (input in $[-1, 1]$). Result ordering from best to worst goes red, orange, yellow. Arrow next to metric indicates whether higher or lower is better.

explains why most of the methods failed on this dataset.

5.1 Dataset Discussion

The limitations of current radiance field methods are most apparent when they are deployed on in-the-wild data that deviates from idealized capture conditions. The San Marino datasets, particularly the "Streets" capture, serve as a critical test case, revealing fundamental weaknesses that standard benchmarks like Mip-NeRF360 fail to expose.

The primary challenge in the "Streets" dataset stems from its capture methodology: a multi-camera rig held at a near-constant height with a predominantly forward-facing trajectory. This motion creates a geometrically degenerate scenario, especially for the ground plane. Without significant changes in camera elevation, the parallax required to triangulate points on the ground is minimal to non-existent. This lack of geometric constraint is catastrophic for most NeRF-based methods. This problem highlights a philosophical divide between classical Structure-from-Motion (SfM) and modern neural radiance fields. In an SfM pipeline, features that cannot be robustly triangulated from multiple viewpoints are simply filtered out and excluded from the sparse reconstruction. Ambiguity leads to data rejection. NeRFs, by contrast, are optimized to explain every pixel from every given camera pose. When faced with geometric ambiguity, the optimization does not filter the region; it often "explodes," leading to the cloudy, detached "floater" artifacts seen in the 3DGS and DBS results. This can be seen in the last row of Fig. 2.

Furthermore, SfM is inherently more robust to the variable lighting conditions and exposure differences present in these datasets. Its reliance on lighting-invariant feature descriptors, such as SIFT (Lowe, 2004), allows it to establish a sparse geometric scaffold even with strong shadows or changes in exposure. The third row of Fig. 2 shows an images with particularly severe exposure difference. The white color needs to come from somewhere, so the model adds a white primitive to the scene that shouldn't be there. While recent work has explored integrating Bayesian filtering and uncertainty estimation into radiance fields to discard these "floater" artifacts (Goli et al., 2024), such techniques are not yet a standard component of mainstream methods. The stark failure of most methods on the "Streets" dataset underscores the need for NeRFs to incorporate more of the geometric rigor and filtering mechanisms that have long been central to classical 3D vision.

5.2 NeRF with Imperfect Data

The nature of large scale multi-scalar mapping is working with imperfect data. It often requires integrating multiple, disparate capture sessions. This process introduces a host of real-world problems that push current methods beyond their limits. During the capture of the San Marino aerial dataset, for instance, the drone's battery life necessitated multiple flights. Between these sessions, the weather changed, altering the global illumination of the entire scene. Another simple problem: many things can't handle different image sizes. Regularizers that could be helpful in dealing with these variations, like MAST3R (Leroy et al., 2024) (Wang et al., 2024a), only accepts image pairs with the same image size.

Current solutions for photometric variation, such as GLO vectors (Duckworth et al., 2024) or Bilateral Grids (Wang et al., 2024b), are insufficient for this task. These methods are designed to model discrepancies in camera hardware (e.g., exposure, white balance), assuming an otherwise static scene. A change in weather, however, is a dynamic change to the scene itself, fundamentally altering how light interacts with every surface. It is less a camera property and more akin to a massive, transient object that cannot be modeled with a simple per-image latent code.

6. Conclusion

This work evaluated current state-of-the-art radiance field methods on challenging real-world datasets, giving insights into their robustness, failure modes, and the limitations of standard evaluation benchmarks. Our findings demonstrate that while methods have matured on idealized datasets like Mip-NeRF360, significant challenges remain for their practical deployment in uncontrolled environments.

Our most important finding is that there can be massive gaps in performance between NeRF methods for challenging datasets. Trajectories with low parallax, such as the ground-level street capture, cause issues for all NeRF methods, but these issues are not simply a matter of fixing the capture. The relative success of EVER shows that it is possible to improve NeRF methods to handle problematic trajectories. The failure of Hierarchical 3DGS also could also potentially be mitigated by using a more robust approach, like EVER, as the backbone rendering method.

Based on our analysis, we offer the following recommendations. For captures that are sparse, large-scale, or feature challenging, non-ideal camera trajectories, EVER is the most suitable choice. Despite its high computational cost and susceptibility to ray termination artifacts, its underlying representation is uniquely capable of preserving structural coherence where other methods fail. For more conventional captures where performance and detail are the priority, SVRaster presents the best balance. It delivers sharp results at interactive frame rates, outperforming the original 3DGS by mitigating the blurriness on distant objects.

7. Future Work

NeRF does not seem quite ready for field work. Further investigation into different ways to accomplish view synthesis could result in improved robustness, which is necessary for field data that is often quite messy.

Another path could lie in utilizing generative models to homogenize data. Recent work has shown that generative video models can be used to create the data necessary to train a model to ”align” data to a specific time, removing the affect of weather (Trevithick et al., 2025). However, this work is still in it’s early stages, and any kind of learning based model can cause problems when the input data is out of distribution.

This new frontier of mapping with imperfect, multi-session data requires its own benchmarks and evaluation protocols. It is inevitable that real-world data will be photometrically and geometrically inconsistent. Yet, as our results indicate, we should not let the perfect be the enemy of the good. Our findings show that perceptual metrics like LPIPS and SSIM are robust enough to guide meaningful progress, even in the face of these challenges. For now, they provide a reliable signal to iterate on methods that can withstand the rigors of real-world data.

Acknowledgements

The authors would like to thank the Republic of San Marino’s Secretary of State for Territory and Environment, and Secretary of State for Education for their support.

References

Atik, M. E., 2025. Comparative Assessment of Neural Radiance Fields and 3D Gaussian Splatting for Point Cloud Generation from UAV Imagery. *Sensors*, 25(10), 2995.

Barron, J. T., Mildenhall, B., Verbin, D., Srinivasan, P. P., Hedman, P., 2022. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5470–5479.

Clini, P., Nespeca, R., Angeloni, R., Coppetta, L., 2024. 3D representation of Architectural Heritage: a comparative analysis of NeRF, Gaussian Splatting, and SfM-MVS reconstructions using low-cost sensors. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 93–99.

Duckworth, D., Hedman, P., Reiser, C., Zhizhin, P., Thibert, J.-F., Lučić, M., Szeliski, R., Barron, J. T., 2024. Smerf: Streamable memory efficient radiance fields for real-time large-scene exploration. *ACM Transactions on Graphics (TOG)*, 43(4), 1–13.

Goli, L., Reading, C., Sellán, S., Jacobson, A., Tagliasacchi, A., 2024. Bayes’ rays: Uncertainty quantification for neural radiance fields. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20061–20070.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4), 1–14.

Kheradmand, S., Rebain, D., Sharma, G., Sun, W., Tseng, Y.-C., Isack, H., Litany, O., Yi, K. M., 2024. 3d gaussian splatting as markov chain monte carlo. M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan (eds), *Advances in Neural Information Processing Systems*, 36, Curran Associates, Inc., 80965–80986.

Knapitsch, A., Park, J., Zhou, Q.-Y., Koltun, V., 2017. Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction. *ACM Transactions on Graphics*, 36(4).

Leroy, V., Cabon, Y., Revaud, J., 2024. Grounding image matching in 3d with mast3r.

Liu, R., Sun, D., Chen, M., Wang, Y., Feng, A., 2025. Deformable Beta Splatting. *arXiv preprint arXiv:2501.18630*.

Lo, E., Meyer, D., Zheng, E., Trinh, S., Forcellini, D., Guerra, G., Cultural Heritage Engineering Initiative (CHEI), 2023. Mount titano - photogrammetry - aerial.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60, 91–110.

Mai, A., Hedman, P., Kopanas, G., Verbin, D., Futschik, D., Xu, Q., Mildenhall, B., Barron, J. T., Zhang, Y., 2024. Ever: Exact volumetric ellipsoid rendering for real-time view synthesis. *arXiv preprint arXiv:2405.01804*.

Martin-Brualla, R., Radwan, N., Sajjadi, M. S. M., Barron, J. T., Dosovitskiy, A., Duckworth, D., 2021. Nerf in the wild: Neural radiance fields for unconstrained photo collections. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7210–7219.

Mazzacca, G., Karami, A., Rigon, S., Farella, E., Trybala, P., Remondino, F. et al., 2023. NeRF for heritage 3D reconstruction. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48(M-2-2023), 1051–1058.

McAvoy, S., Forcellini, D., G., G., Kuester, F., 2024. Mt. Titano Historic Center Street Survey, San Marino. *OpenHeritage3D*.

McAvoy, S., Forcellini, D., Kuester, F., 2025. San Marino - Historic City Center Earthquake Resilience Survey March 4th 2024. *eScholarship*.

McAvoy, S., Ristevski, J., Rissolo, D., Kuester, F., 2023. OPEN-HERITAGE3D: BUILDING AN OPEN VISUAL ARCHIVE FOR SITE SCALE GIGA-RESOLUTION LIDAR AND PHOTOGRAMMETRY DATA. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-M-1-2023, 215–222.

Meyer, D., Lo, E., Klingspon, J., Netchaev, A., Ellison, C., Kuester, F., 2020. TunnelCAM- A HDR Spherical Camera Array for Structural Integrity Assessments of Dam Interiors. *Electronic Imaging*, 32number 7, Society for Imaging Science and Technology, 227–1–227–7.

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.

Schonberger, J. L., Frahm, J.-M., 2016. Structure-from-motion revisited. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4104–4113.

Sun, C., Choe, J., Loop, C., Ma, W.-C., Wang, Y. C. F., 2024. Sparse Voxels Rasterization: Real-time High-fidelity Radiance Field Rendering. *arXiv preprint arXiv:2405.04459*.

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. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8248–8258.

Trevithick, A., Paiss, R., Henzler, P., Verbin, D., Wu, R., Alzayer, H., Gao, R., Poole, B., Barron, J. T., Holynski, A. et al., 2025. Simvs: Simulating world inconsistencies for robust view synthesis. *Proceedings of the Computer Vision and Pattern Recognition Conference*, 16464–16474.

Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., Revaud, J., 2024a. Dust3r: Geometric 3d vision made easy. *CVPR*.

Wang, Y., Wang, C., Gong, B., Xue, T., 2024b. Bilateral Guided Radiance Field Processing. *ACM Transactions on Graphics (TOG)*, 43(4), 1–13.

Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600–612.

Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H., 2024. Depth anything v2. *Advances in Neural Information Processing Systems*, 37, 21875–21911.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. *CVPR*.