# MACGaussian: Robust 3D Gaussian Splatting from Sparse Input Views Using High-Precision Measurement-Arm-Camera (MAC) Capture

Saptarshi Neil Sinha[1(✉)], Muhammad Ali Shahid[1], and Michael Weinmann[2]

[1] Fraunhofer IGD, Darmstadt, Germany
saptarshi.neil.sinha@igd.fraunhofer.de
[2] Delft University of Technology, Delft, Netherlands

**Abstract.** Recent techniques like neural radiance fields (NeRFs) and 3D Gaussian splatting (3DGS) have led to significant improvements in novel view synthesis. Whereas the explicit scene representation of 3DGS in terms of Gaussians allows real-time rendering with state-of-the-art quality, this approach relies on the availability of many views to achieve a coherent scene representation. In this paper, we investigate the importance of accurate camera poses and demonstrate that this even allows for accurate scene representation based on 3D Gaussian Splatting in a sparse-view setting. For this purpose, we address accurate pose estimation by employing a measurement arm equipped with a camera, achieving precise camera-pose estimates with sub-millimeter accuracy. Based on a newly introduced dataset (Core dataset) with its accurate pose information, we demonstrate superior quality in terms of quality of rendered novel views in comparison to results achieved based on calibrations with Dust3R-based and COLMAP-based initializations of the 3D Gaussians. Thereby, our approach offers a reliable and effective solution to practical, sparse-view reconstruction for the preservation of cultural heritage artifacts, which is particularly relevant in applications like virtual museums and archaeology. Furthermore, we expect our Core dataset to serve as a reasonable benchmark, advancing the understanding and development of robust 3D reconstruction methods.

**Keywords:** 3D Reconstruction · Gaussian splatting · Novel view synthesis · Pose estimation · Digital restoration · Digital twins

## 1 Introduction

*Digital twins* play a crucial role in preserving and studying cultural heritage by offering precise representations and enabling remote accessibility [16]. This

S. N. Sinha and M. A. Shahid—Equal contribution.


A. Del Bue et al. (Eds.): ECCV 2024 Workshops, LNCS 15627, pp. 235–248, 2025.
https://doi.org/10.1007/978-3-031-92808-6_15

technology has significant applications in fields such as virtual museums, archaeology, paleontology, and physical anthropology, enhancing processes and facilitating research, education, and preservation efforts. Recent advancements in AI methods have played a crucial role in creating and analyzing digital twins, making the process more efficient and inclusive for non-specialists. This combination of digital twins, AI, and user-friendly technologies addresses reproducibility and enables the rapid creation of site-specific digital twins to preserve endangered cultural artifacts and sites. *Digital restoration* using 3D reconstruction enables the virtual restoration of cultural relics, such as the Bodhidharma polychrome sculpture from the Lingyan Temple in China [40], by combining scientific analyses, simulation experiments, 3D scanning, and multi-view 3D reconstruction, offering a new perspective for archaeology, art history, and cultural heritage research.

In cultural heritage applications, evaluating reconstruction techniques is crucial for ensuring reliable and effective 3D models. To overcome challenges like specularity and lack of distinctive features, researchers have developed lighting models and reconstruction techniques based on differentiable rendering [19]. However, when creating digital twins of human remains, various 3D technologies like Computed tomography (CT), laser scanners, and photogrammetry are commonly employed. While CT provides complete volume capture, it can be expensive and inaccessible. Photogrammetry, particularly Structure-from-Motion, is more affordable and portable but may lack accuracy compared to laser scanner-derived models. In recent years, learning-based scene representation and rendering methods such as neural radiance fields [27] and 3D Gaussian splatting [21] have gained a lot of attention. In particular, 3D Gaussian splatting has been demonstrated to allow accurate scene representation and visualization in real time, while also facilitating interpretability and editing due to the underlying explicit scene representation in terms of 3D Gaussians. The number and arrangement of these Gaussians is optimized by incorporating an imaging model, that allows synthesizing images from the underlying Gaussian representation in 3D space, in the optimization of the Gaussians with the goal of matching the appearance in the input images with the images synthesized under the respective camera poses. Combining 3D Gaussian splatting with an initialization based on DUST3R [46] even allowed accurate scene representation for scenarios with limited viewpoints [10].

In this paper, we aim at investigating the importance of accurate camera poses for accurate scene representation based on 3D Gaussian Splatting in a sparse-view setting. To address the challenge of accurate pose estimation, our approach incorporates a measurement arm equipped with a camera [22], providing pose measurements with sub-millimeter accuracy. We achieve this using Hand-Eye-Calibration which allows the camera-poses to be derived from the measurement-arm (measurement-tip) poses. Utilizing this setup, we generate 3D reconstructions for our newly introduced cultural heritage statue dataset, referred to as the Core dataset [20], and demonstrate the potential of our approach in comparison to the alternatives of using other learning-based

initializations like DUST3R [46] or standard Structure-from-Motion (SfM) based initialization as obtained by COLMAP [35]. In particular, our method demonstrates superior performance by achieving higher accuracy and better alignment of reconstructed models compared to DUST3R and COLMAP. This advancement in accurate pose estimation, facilitated by the integration of a measurement arm and calibrated camera, contributes to the field of cultural heritage research and preservation, providing more reliable digital representations of cultural artifacts and statues.

In summary, our main contributions are as follows:

- We present an accurately calibrated measurement-arm-camera (MAC) setup to accurately generate splats using 3D Gaussian Splatting (3DGS), thereby enabling improved sparse-view 3D reconstruction in various applications, including cultural heritage preservation and research.
- We demonstrate the potential of our approach with respect to other alternatives such as the initialization of 3D Gaussian splatting based on DUST3R [46] and COLMAP [35], where we show our approach to achieve higher accuracy and a better alignment of the reconstructed model.
- We introduce the Core dataset, which provides accurate poses and has the potential to significantly advance the research community's understanding and improvement of 3D reconstruction methods in cultural heritage applications. By using this benchmark dataset, researchers can evaluate their algorithms and we will make this dataset available on our website.

## 2 Related Work

In this section, we briefly review developments on radiance based scene representation and novel view generation as well as 6D pose estimation for recent learning-based approaches (i.e., NeRFs and 3DGS).

### 2.1 Radiance Based Scene Representation and Novel View Generation

The Neural radiance fields (NeRF) approach [27] represents the scene using a neural network that predicts local density and view-dependent color for points in the scene volume. These predictions are used to synthesize images through volume rendering techniques. The network is trained by optimizing the predicted images to match the given input images, refining its representation of the scene. Extensions of NeRF address rendering quality issues like aliasing [2–4,44] and accelerate network training [7,12,28,32,53]. There are also efforts to handle complex inputs such as unconstrained image collections [8,18,25], camera pose refinement [17,23,48,55], deformable scenes [30,31], and large-scale scenarios [26,38,42]. Techniques using depth cues have been developed to guide training and handle textureless regions [1,9,33,34,49].

Alternative approaches to scene representation include implicit surfaces [14,45,47], explicit representations using points [51], meshes [29], and 3D Gaussians [21]. Point-based neural rendering techniques like Point-NeRF [51] combine the precise view synthesis of NeRF with the fast scene reconstruction capabilities of multi-view stereo methods. These techniques utilize neural 3D point clouds to enable efficient rendering and accelerated training. Point-based methods have also shown promise for scene editing [56]. A recent state-of-the-art method called 3D Gaussian Splatting (3DGS) [21] surpasses existing implicit neural representation methods in both quality and efficiency. It employs anisotropic 3D Gaussians as an explicit scene representation and a fast differentiable rasterizer for image rendering.

### 2.2 6D Camera Pose Estimation

The practical application of NeRFs and 3DGS is limited due to the requirement of densely captured images and heavy reliance on COLMAP [35,36] for initially providing poses, which demands users with photography expertise and significant computing resources. Furthermore, SfM errors can impact the quality of the 3D representation, leading to potential failures when images lack sufficient overlap and detailed textures. To overcome this challenge, researchers have introduced regularization techniques to optimize the radiance fields and address the need for an adequate number of views.

Structure-from-Motion (SfM) [35] is a technique used to reconstruct sparse 3D maps from a set of images while simultaneously determining camera parameters. The traditional pipeline involves keypoint matching [5,6] and bundle adjustment, but recent enhancements have incorporated learning-based techniques to improve feature description, image matching, feature refinement, and bundle adjustment [24,39,43,50]. However, the sequential nature of the SfM pipeline can still be susceptible to noise and errors in each component that may influence the subsequent steps. Furthermore, the recently presented DUSt3R [46] operates without prior information about camera calibration or viewpoint poses, utilizing a regression-based approach for end-to-end 3D reconstruction, unifying monocular and binocular reconstruction cases, and employing a global alignment strategy. Besides leveraging a powerful pre-trained model trained on several indoor and outdoor datasets, DUST3R involves a sophisticated architecture based on transformer encoders and decoders, resulting in state-of-the-art performance in depth estimation and relative pose estimation.

Once camera parameters have been computed, dense scene reconstruction can be approached based on standard Multi-View Stereo (MVS) approaches [37, 52,54] or the aforementioned learning-based approaches based on NeRFs [27] and 3DGS [21]. While COLMAP serves as a standard initialization for 3DGS [21], InstantSplat [10] leverages an initialization based on DUST3R [46] to enable rapid scene reconstruction from sparse-view, unposed images. However, it should be noted that InstantSplat [10] has limitations in scenes with a large number of images, as it requires globally aligned point clouds.

To address the limitations of multi-view stereo and structure-from-motion-based pose estimation, we introduce a measurement-arm-camera (MAC) setup [11] which leverages the highly accurate poses delivered by the measurement-arm and a hand-eye calibration [41] to provide accurate camera-poses (Fig. 2). This calibrated setup with accurate camera-poses can be utilized for improved multi-view reconstruction using 3DGS. Both metrology (measurement-arm) and (hand-eye) calibration bring a high-degree of certainty to the camera-pose estimation task and therefore help overcome the challenges associated with traditional pose estimation techniques. Thus, the accuracy and reliability of the multi-view reconstruction process can be enhanced greatly.

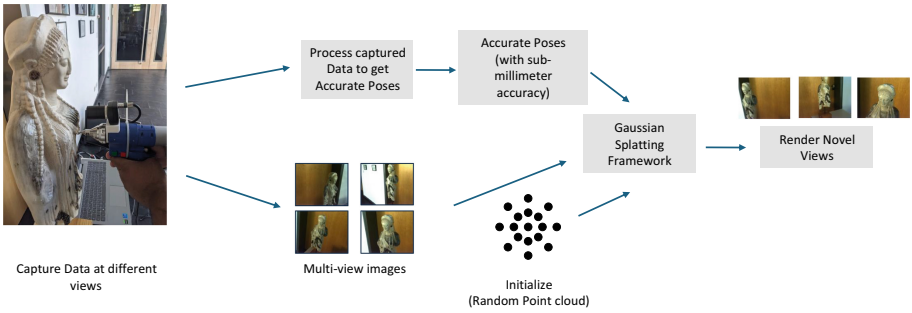



**Fig. 1.** Pipeline for capturing and generating novel views using Measurement-arm-camera (MAC) and 3D Gaussian splatting

## 3 Methodology

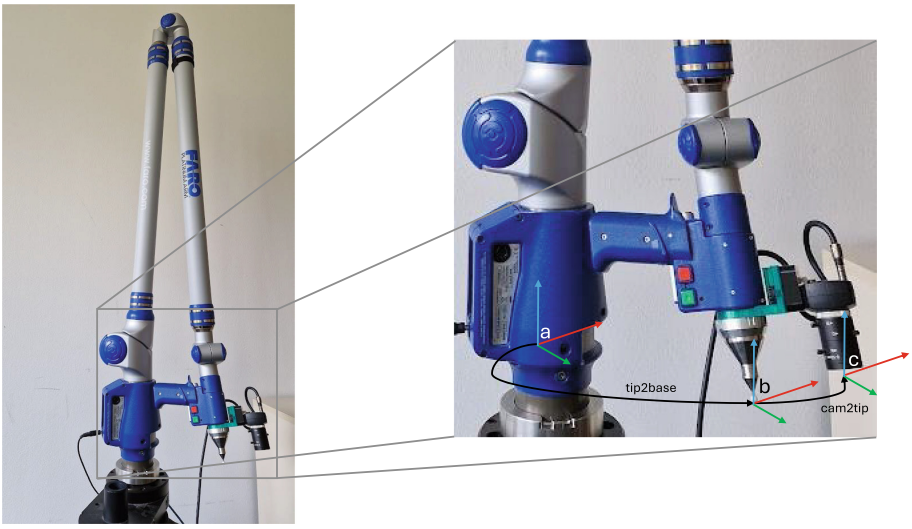
In this section, we present a pipeline (Fig. 1) that involves the development of a measurement-arm-camera setup to accurately capture datasets for generating splats using a 3D Gaussian splatting framework. Using our setup, we capture highly accurate poses from different viewpoints along with corresponding images from the attached camera. This data is then processed and converted into a format compatible with the 3D Gaussian Splatting (3DGS) pipeline for training.

### 3.1 Measurement-Camera-Arm

Our Measurement-Arm-Camera setup consists of a Faro Platinum (P12 7-Axis) measurement arm with an accuracy rating (single-point-accuracy) of 0.073 mm. The camera is a *uEye UI324xML-C* camera with an image resolution of 1280 × 1024 pixels. The camera is attached to the measurement arm tip using a specially printed adapter. We employ a Hand-Eye-Calibration tool-chain which processes calibration data in the following steps:

– Synchronization of the measurement arm tip and the camera poses (obtained using Charuco-Board [13] tracking)
– Calculation of initial Hand-Eye-Calibration.
– Outlier removal.
– Calculation of final Hand-Eye-Calibration. (The Hand-Eye-Calibration is done using [15])

The final calibration has residuals in the sub-millimeter (translation component) and sub-degree (rotation component) ranges.




a) Measurement-Arm Base (considered world origin) b) Measurement-Tip c) Camera Coordinate-System

tip2base: The pose of the Measurement-Tip in the Measurement-Arm Base Coordinate-System.
cam2tip: The pose of the Camera in the Measurement-Tip Coordinate-System obtained using the Hand-Eye-Calibration.

**Fig. 2.** Measurement-Arm-Camera setup showing the different coordinate-systems and transformations.

Using this calibration, a highly-accurate estimate of the pose of the camera relative to the measurement-arm tip is obtained. This combined with the extremely accurate measurement arm tip pose obtained from the measurement arm, allows a very high-precision estimate of the camera pose (relative to the measurement arm) to be determined during data capture. A slight inaccuracy can be introduced by the asynchronous data streams of the two devices, but this is eliminated by capturing data-pairs (measurement arm tip pose and camera image) from the devices while the camera-tip combination is held still.

Thus the data-capture process consists of capturing snapshots of the measurement tip and camera as pose-image pairs from different locations around the object of interest, by holding the camera still at these locations and capturing a snapshot by pressing a button on the measurement arm grip.

The captured data is then processed to calculate the camera poses (using the pose of the measurement arm tip and the Hand-Eye-Calibration) and these are stored in a COLMAP format, to be loaded in the Gaussian splatting pipeline and used to train the 3D Gaussian Splats. The camera intrinsics calculated to use during the camera-pose estimation part of the Hand-Eye-Calibration are also provided in the same COLMAP format to the Gaussian Splatting pipeline.

### 3.2 Gaussian Splatting

After data creation and initialization using our MAC setup, we follow the original 3D Gaussian Splatting formulation [21] for scene representation and rendering. Training is conducted with default parameters and learning rates of 0.0025 for spherical harmonics features, 0.05 for opacity adjustments, 0.005 for scaling operations, and 0.001 for rotation transformations. The training procedure consists of 30,000 iterations using an NVIDIA RTX3090 GPU. The total loss for optimization is defined by the following Eq. 1 as in the original implementation:

$$L = (1 - \lambda)L_1 + \lambda L_{D-SSIM} \tag{1}$$

Here, $\lambda$ is set to 0.2 by default. $L_1$ represents the L1-Norm of the per-pixel color difference, and $L_{D-SSIM}$ represents the $L_{D-SSIM}$ term [21].

**Fig. 3.** Core dataset [20] with different view directions

## 4 Experiments

We perform both quantitative analysis (PSNR, SSIM and LPIPS) and qualitative analysis on our novel Core dataset as well as comparisons to baseline approaches to evaluate the potential of our methodology.

### 4.1 Dataset

Using a measurement camera arm setup, we generated a novel dataset (Core dataset) by capturing images and corresponding poses, with a snapshot of some views shown in Fig. 3. We created a sparse dataset with 11 views with accurate poses.

### 4.2 Competing Baseline Methods

To demonstrate the potential of our approach, we perform comparisons to two different baseline approaches.

As the first baseline, we use conventional Structure-from-Motion (SfM) [35] as implemented in COLMAP and used for the initialization of the original 3D Gaussian Splatting approach [21]. SfM calculates and matches point correspondences in overlapping images, resulting in a sparse 3D reconstruction of the scene and camera parameters (including camera poses and intrinsics). In the initialization of Gaussian Splatting, the respective camera parameters are associated with the respective images and the sparse point cloud is used as initialization for the 3D Gaussians, which are afterwards adjusted in their number, shape and arrangement during optimization to create a dense reconstruction.

As a second baseline, we use DUST3R [46], a recent learning-based end-to-end approach for 3D scene reconstruction. This is motivated by the fact that the emergence of deep learning-based dense stereo frameworks has revolutionized depth map prediction, significantly reducing processing times to milliseconds and, hence, offering improved efficiency over the time-consuming Structure-from-Motion (SfM) process and offering remarkable improvements in accelerating stereo matching. DUSt3R [46] employs an efficient optimization technique to solve for dense per-view pixel-to-3D mappings that can be used for camera calibration, depth estimation and dense 3D reconstruction, while also providing confidence maps to indicate the reliability of the reconstruction. The approach involves a sophisticated transformer-based network architecture that has been trained based on multiple indoor and outdoor datasets.

### 4.3 Quantitative Analysis

The quantitative analysis is presented in Table 1. The table presents quantitative comparisons (PSNR, SSIM, LPIPS) for different numbers of views used for scene reconstruction, comparing the performance of DUST3R, COLMAP, and our approach. Our method consistently outperforms the other two baselines in terms of PSNR and SSIM, indicating higher image quality and a higher similarity to ground truth. Additionally, our approach achieves the lowest LPIPS scores, indicating better perceptual similarity to the reference images.

### 4.4 Qualitative Analysis

During our qualitative analysis, we compared the pixel differences between the rendered images and ground-truth images using heat-maps. Figure 4 demonstrates the differences observed when all captured sparse-views were utilized,

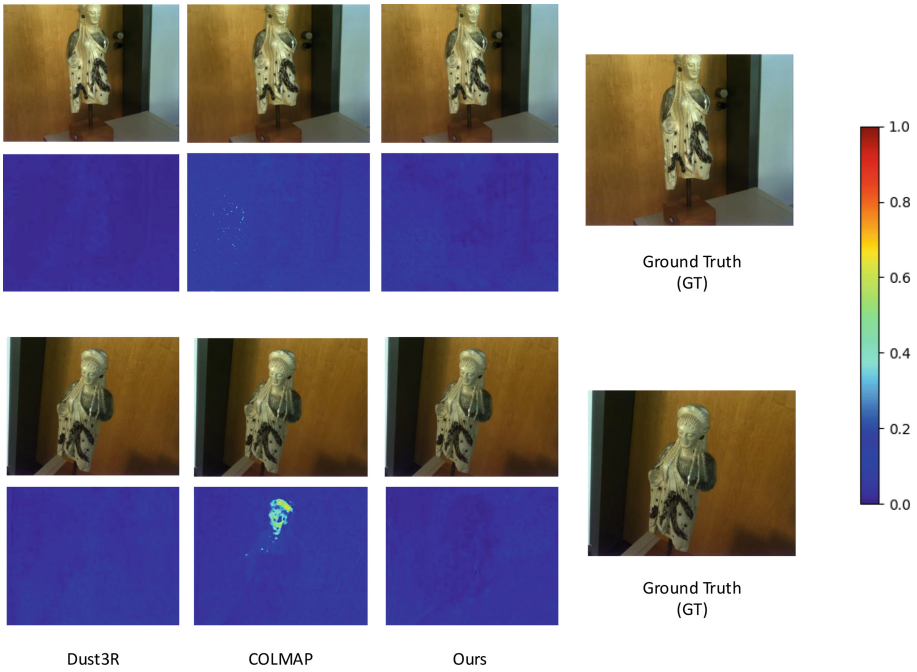



**Fig. 4.** Qualitative analysis with two different test poses on the training result trained on all multi-view images

**Table 1.** Quantitative Comparisons (PSNR / SSIM / LPIPS) for different number of views used for scene reconstruction.

| Number of views | Dust3R | COLMAP | Ours |
|---|---|---|---|
| | PSNR ↑ | | |
| 11 | 40.055 | 39.708 | 41.418 |
| 8 | 40.789 | 41.433 | 42.415 |
| 6 | 40.929 | - | 41.989 |
| 3 | 40.386 | - | 40.395 |
| | SSIM ↑ | | |
| 11 | 0.950 | 0.954 | 0.964 |
| 8 | 0.956 | 0.964 | 0.969 |
| 6 | 0.956 | - | 0.969 |
| 3 | 0.953 | - | 0.954 |
| | LPIPS ↓ | | |
| 11 | 0.195 | 0.223 | 0.155 |
| 8 | 0.168 | 0.185 | 0.128 |
| 6 | 0.164 | - | 0.124 |
| 3 | 0.190 | - | 0.195 |

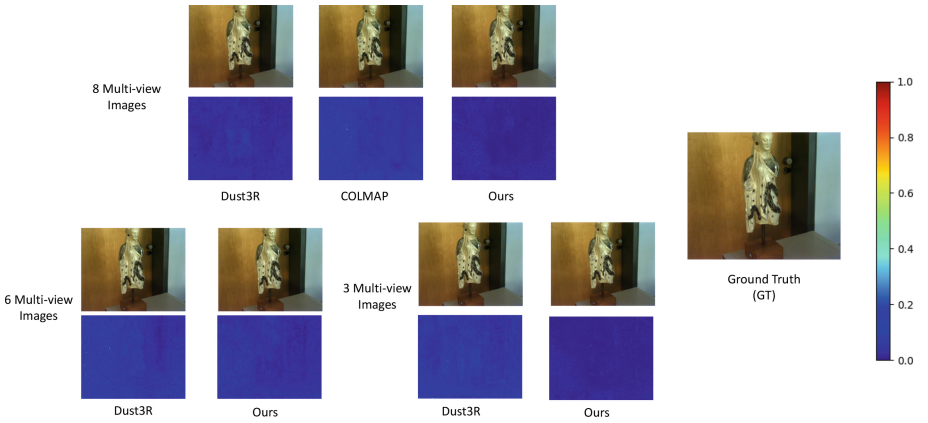



**Fig. 5.** Qualitative analysis on the training result trained on different multi-view images

while Fig. 5 illustrates the variations when using different numbers of views (8, 6, and 3). The heat-maps clearly indicate that both DUST3R and our method outperform COLMAP in terms of rendering using 3DGS. Additionally, our method shows improved reconstruction quality when compared with DUST3R, particularly with fewer views, as evident from the heat-map analysis.

## 5 Limitations

The limitations of our method include the need to improve the initialization process of the point cloud to achieve better convergence. Currently, the initialization of 3D Gaussian splatting with COLMAP leads to faster convergence compared to our method and Dust3R. However, it is important to note that the rendered outputs is not as accurate after completing the full training of 30000 iterations. Furthermore, our evaluation did not include a comprehensive analysis of the results with the different methods for a large number of views, where COLMAP typically performs better in reconstructing poses than for the sparse-view scenario. Nevertheless, our method specifically addresses the challenge of 3D reconstruction from sparse views in cultural heritage applications, where limited availability of past data is a significant constraint.

## 6 Conclusion

In conclusion, this study presents a novel approach for 3D reconstruction in cultural heritage preservation. By integrating a measurement arm with a calibrated camera (MAC setup) we provide an approach to high-quality initialization of 3D Gaussian Splatting (3D GS) improving over pose estimation accuracy and model alignment accuracy provided by existing methods like COLMAP [35] and DUST3R [46]. Thereby, our approach also improves the accuracy of scene

representation and rendering based on 3D Gaussian Splatting [21]. Based on both qualitative and quantitative evaluation, we validate the effectiveness of our method, where we confirm improved reconstruction quality, even with fewer views. The introduction of the Core dataset, with accurate poses, contributes to advancing 3D reconstruction methods in cultural heritage applications. This research addresses current challenges in 3D reconstruction and establishes a foundation for reliable and efficient cultural heritage preservation and analysis. Future work includes real-time reconstruction and optimizing view selection for accurate geometry reconstruction using the free capture of the measurement arm.

**Acknowledgement.** The work in this paper was partially funded by the European Commission for the PERCEIVE project (grant agreement 101061157).

## References

1. Attal, B., et al.: TöRF: time-of-flight radiance fields for dynamic scene view synthesis. NeurIPS **34**, 26289–26301 (2021)
2. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields (2021)
3. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-NeRF 360: unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5460–5469 (2022)
4. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-NeRF: anti-aliased grid-based neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 19697–19705 (2023)
5. Barroso-Laguna, A., Riba, E., Ponsa, D., Mikolajczyk, K.: Key.Net: Keypoint detection by handcrafted and learned CNN filters. In: ICCV, pp. 5836–5844 (2019)
6. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: speeded up robust features. In: ECCV, pp. 404–417. Springer (2006)
7. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: TensoRF: tensorial radiance fields. In: Proceedings of the European Conference on Computer Vision, pp. 333–350 (2022)
8. Chen, X., et al.: Hallucinated neural radiance fields in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12943–12952 (2022)
9. Deng, K., Liu, A., Zhu, J.Y., Ramanan, D.: Depth-supervised NeRF: fewer views and faster training for free. In: CVPR, pp. 12882–12891. IEEE (2022). https://doi.org/10.1109/cvpr52688.2022.01254
10. Fan, Z., et al.: Instantsplat: unbounded sparse-view pose-free gaussian splatting in 40 seconds. arXiv preprint arXiv:2403.20309 (2024)
11. FARO: Understanding portable measurement arms. https://www.faro.com/en/Resource-Library/Article/understanding-portable-measurement-arms. Accessed 31 July 2024
12. Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: radiance fields without neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5491–5500 (2022)

13. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Marín-Jiménez, M.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recogn. **47**(6), 2280–2292 (2014). https://doi.org/10.1016/j.patcog.2014.01.005
14. Ge, W., Hu, T., Zhao, H., Liu, S., Chen, Y.C.: Ref-NeuS: ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection. arXiv preprint arXiv:2303.10840 (2023)
15. Horaud, R., Dornaika, F.: Hand-eye calibration. Int. J. Robot. Res. **14**(3), 195–210 (1995)
16. Hutson, J., Weber, J., Russo, A.: Digital twins and cultural heritage preservation: a case study of best practices and reproducibility in Chiesa dei ss Apostoli e Biagio. Art Design Res. **11**(1) (2023). https://doi.org/10.4236/adr.2023.111003
17. Jeong, Y., Ahn, S., Choy, C., Anandkumar, A., Cho, M., Park, J.: Self-calibrating neural radiance fields. In: ICCV, pp. 5846–5854. IEEE (2021). https://doi.org/10.1109/iccv48922.2021.00579
18. Jun-Seong, K., Yu-Ji, K., Ye-Bin, M., Oh, T.H.: HDR-Plenoxels: self-calibrating high dynamic range radiance fields. In: Proceedings of the European Conference on Computer Vision, pp. 384–401 (2022)
19. Kato, H., et al.: Differentiable rendering: a survey. arXiv preprint arXiv:2006.12057 (2020)
20. Keil, J., et al.: A digital look at physical museum exhibits: designing personalized stories with handheld augmented reality in museums. In: 2013 Digital Heritage International Congress (DigitalHeritage), vol. 2, pp. 685–688 (2013). https://doi.org/10.1109/DigitalHeritage.2013.6744836
21. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4) (2023). https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/
22. Kovač, I., Frank, A.: Testing and calibration of coordinate measuring arms. Precision Eng. **25**(2), 90–99 (2001). https://doi.org/10.1016/S0141-6359(00)00057-X
23. Lin, C.H., Ma, W.C., Torralba, A., Lucey, S.: BaRF: bundle-adjusting neural radiance fields. In: ICCV, pp. 5741–5751. IEEE (2021). https://doi.org/10.1109/iccv48922.2021.00569
24. Lindenberger, P., Sarlin, P.E., Larsson, V., Pollefeys, M.: Pixel-perfect structure-from-motion with featuremetric refinement. In: ICCV (2021)
25. Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D.: NeRF in the wild: neural radiance fields for unconstrained photo collections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7206–7215 (2021)
26. Mi, Z., Xu, D.: Switch-NeRF: learning scene decomposition with mixture of experts for large-scale neural radiance fields. In: ICLR (2023)
27. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
28. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. **41**(4), 102:1–102:15 (2022)
29. Munkberg, J., et al.: Extracting triangular 3D models, materials, and lighting from images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8270–8280 (2022)
30. Park, K., et al.: Nerfies: deformable neural radiance fields. In: ICCV, pp. 5865–5874. IEEE (2021). https://doi.org/10.1109/iccv48922.2021.00581

31. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-NeRF: neural radiance fields for dynamic scenes. In: CVPR, pp. 10318–10327. IEEE (2021). https://doi.org/10.1109/cvpr46437.2021.01018
32. Reiser, C., Peng, S., Liao, Y., Geiger, A.: Kilonerf: speeding up neural radiance fields with thousands of tiny MLPs. In: ICCV, pp. 14335–14345 (2021)
33. Rematas, K., et al.: Urban radiance fields. In: CVPR, pp. 12932–12942. IEEE (2022). https://doi.org/10.1109/cvpr52688.2022.01259
34. Roessle, B., Barron, J.T., Mildenhall, B., Srinivasan, P.P., Nießner, M.: Dense depth priors for neural radiance fields from sparse input views. In: CVPR, pp. 12892–12901. IEEE (2022). https://doi.org/10.1109/cvpr52688.2022.01255
35. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
36. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV) (2016)
37. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (2016)
38. Tancik, M., et al.: Block-NeRF: scalable large scene neural view synthesis. In: CVPR, pp. 8248–8258. IEEE (2022). https://doi.org/10.1109/cvpr52688.2022.00807
39. Tang, S., Zhang, J., Zhu, S., Tan, P.: Quadtree attention for vision transformers. In: ICLR (2022)
40. Tong, Y., Cai, Y., Nevin, A., Ma, Q.: Digital technology virtual restoration of the colours and textures of polychrome Bodhidharma statue from the Lingyan temple, Shandong. China. Herit. Sci. **11**(1), 12 (2023). https://doi.org/10.1186/s40494-023-00858-y
41. Tsai, R., Lenz, R.: A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. IEEE Trans. Robot. Autom. **5**(3), 345–358 (1989)
42. Turki, H., Ramanan, D., Satyanarayanan, M.: Mega-NeRF: scalable construction of large-scale NeRFs for virtual fly-throughs. In: CVPR, pp. 12922–12931. IEEE (2022). https://doi.org/10.1109/cvpr52688.2022.01258
43. Tyszkiewicz, M., Fua, P., Trulls, E.: Disk: learning local features with policy gradient. In: Advances in Neural Information Processing Systems, vol. 33, pp. 14254–14265 (2020)
44. Wang, C., Wu, X., Guo, Y.C., Zhang, S.H., Tai, Y.W., Hu, S.M.: NeRF-SR: high quality neural radiance fields using supersampling. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 6445–6454 (2022)
45. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction. Adv. Neural Inf. Proc. Syst. **354**, 27171–27183 (2021)
46. Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., Revaud, J.: DUSt3R: geometric 3D vision made easy. In: CVPR (2024)
47. Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L.: Neus2: fast learning of neural implicit surfaces for multi-view reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3272–3283 (2023)
48. Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V.A.: NeRF–: neural radiance fields without known camera parameters (2021)

49. Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J., Zhou, J.: NerfingMVS: guided optimization of neural radiance fields for indoor multi-view stereo. In: ICCV, pp. 5610–5619. IEEE (2021). https://doi.org/10.1109/iccv48922.2021.00556
50. Xiao, L., Xue, N., Wu, T., Xia, G.S.: Level-s2fm: structure from motion on neural level set of implicit surfaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
51. Xu, Q., et al.: Point-NeRF: point-based neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5438–5448 (2022)
52. Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L.: MVSNet: depth inference for unstructured multi-view stereo. In: European Conference on Computer Vision (2018)
53. Yariv, L., et al.: BakedSDF: meshing neural SDFs for real-time view synthesis. In: Brunvand, E., Sheffer, A., Wimmer, M. (eds.) Proceedings of the ACM SIGGRAPH 2023 Conference, pp. 46:1–46:9 (2023)
54. Ye, X., Zhao, W., Liu, T., Huang, Z., Cao, Z., Li, X.: Constraining depth map geometry for multi-view stereo: a dual-depth approach with saddle-shaped depth cells. In: International Conference on Computer Vision (2023)
55. Yen-Chen, L., Florence, P., Barron, J.T., Rodriguez, A., Isola, P., Lin, T.Y.: iNeRF: inverting neural radiance fields for pose estimation, pp. 1323–1330. IEEE (2021). https://doi.org/10.1109/iros51168.2021.9636708
56. Zhang, Y., Huang, X., Ni, B., Li, T., Zhang, W.: Frequency-modulated point cloud rendering with easy editing. arXiv preprint arXiv:2303.07596 (2023)