





TRIPS: Trilinear Point Splatting for Real-Time Radiance Field Rendering

Linus Franke , Darius Rückert , Laura Fink  and Marc Stamminger 

Visual Computing Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
{firstname.lastname}@fau.de



Figure 1: Previous point-based radiance field rendering methods provide great results in many cases, but renderings can be aliased and incomplete (ADOP [RFS22] (left), missing parts of the bike’s tire), or overblurred (3D Gaussian splatting [KKLD23] (middle), missing fine grass details). Our approach combines the advantages of both to render crisp, complete, and alias-free images.

Abstract

Point-based radiance field rendering has demonstrated impressive results for novel view synthesis, offering a compelling blend of rendering quality and computational efficiency. However, also latest approaches in this domain are not without their shortcomings. 3D Gaussian Splatting [KKLD23] struggles when tasked with rendering highly detailed scenes, due to blurring and cloudy artifacts. On the other hand, ADOP [RFS22] can accommodate crisper images, but the neural reconstruction network decreases performance, it grapples with temporal instability and it is unable to effectively address large gaps in the point cloud. In this paper, we present TRIPS (Trilinear Point Splatting), an approach that combines ideas from both Gaussian Splatting and ADOP. The fundamental concept behind our novel technique involves rasterizing points into a screen-space image pyramid, with the selection of the pyramid layer determined by the projected point size. This approach allows rendering arbitrarily large points using a single trilinear write. A lightweight neural network is then used to reconstruct a hole-free image including detail beyond splat resolution. Importantly, our render pipeline is entirely differentiable, allowing for automatic optimization of both point sizes and positions.

Our evaluation demonstrate that TRIPS surpasses existing state-of-the-art methods in terms of rendering quality while maintaining a real-time frame rate of 60 frames per second on readily available hardware. This performance extends to challenging scenarios, such as scenes featuring intricate geometry, expansive landscapes, and auto-exposed footage.

The project page is located at: <https://lfranke.github.io/trips>

CCS Concepts

• **Computing methodologies** → **Rendering; Image-based rendering; Reconstruction;**

1. Introduction

Novel view synthesis methods have been a significant driver for computer graphics and vision, as they have revolutionized the way we perceive and interact with 3D scenes. Many of these methods rely on explicit representations, such as meshes or points. Typically, the explicit models are derived from 3D reconstruction processes and can be efficiently rendered through rasterization, which aligns well with contemporary GPU capabilities. Nevertheless, these reconstructed models often fall short of perfection and necessitate additional steps to mitigate artifacts.

A common strategy to handle these artifacts is to use scene-specific optimization methods, known as inverse rendering. This allows for the adjustment of the scene's texture, geometry, and camera parameters to align the rendering with the photograph. Prominent techniques in this domain incorporate per-point descriptors [RFS22, ASK*20, KPLD21], explicit optimization of point sizes via Gaussians [KKLD23] and learned neural refinement networks [TZN19, RFS22, KLR*22]. While this generally extends render times, it significantly enhances visual quality.

In the realm of point-based inverse and neural rendering techniques, two successful recent approaches are *3D Gaussian Splatting* [KKLD23] and *ADOP* [RFS22]. The former method employs a unique strategy where each point is rendered as a 3D Gaussian distribution, allowing for direct optimization of the points' shape and size. This process effectively fills gaps in point clouds within the global coordinate space through the utilization of large splats. Remarkably, this approach yields high-quality images without necessitating the integration of a neural network for reconstruction. However, a drawback is the potential loss of sharpness, as Gaussians tend to introduce blurriness and cloudy artifacts, particularly when there are limited observations available.

In contrast, ADOP rasterizes radiance fields as one-pixel points with depth testing at multiple resolutions. Subsequently, it employs a neural network to address gaps and enhance texture details in screen space. This approach possesses the capability to reconstruct texture details that surpass the resolution of the original point cloud, although the neural network adds an additional computational overhead and shows weaknesses in filling large holes.

In this paper, we introduce TRIPS, a novel approach that seeks to harness the strengths of both ADOP and 3D Gaussians without losing real-time rendering capabilities. Similar to 3D Gaussian Splatting, TRIPS rasterizes splats of varying size, however, like ADOP, it also applies a reconstruction network to generate hole-free and crisp images. More precisely, we first rasterize the point cloud as $2 \times 2 \times 2$ trilinear splats into an image pyramid and blend them using front-to-back alpha blending. Subsequently, we feed the image pyramid through a compact and efficient neural reconstruction network, which harmonizes the various layers, addresses remaining gaps, and conceals rendering artifacts. To ensure the preservation of high levels of detail, particularly in challenging input scenarios, we incorporate spherical harmonics and a tone mapping module into our pipeline.

In our evaluations, we demonstrate that our approach can yield crisper images compared to 3D Gaussians, with almost the same performance. Furthermore, it surpasses ADOP in the task of filling

sizable gaps and maintaining temporal consistency throughout the rendering process. In summary, our contributions are:

- The introduction of TRIPS, a novel trilinear point splatting technique for radiance field rendering.
- A differentiable pipeline for optimization of all input parameters, including point positions and sizes, creating a robust scene representations.
- An implementation of the method resulting in high-quality real-time renderings under varying capturing conditions at:

<https://github.com/lfranke/TRIPS>

2. Related Work

In this section, we provide an overview of the field of novel view synthesis and choices for scene representations in this problem domain.

2.1. Novel View Synthesis and Traditional Approaches

Traditionally, real-world novel view synthesis relies on image-based rendering techniques. Commonly, *Structure-from-Motion (SfM)* techniques [SSS06, SF16] allow camera parameter estimations from a set of photographs which are then used for directly warping source image colors to a target view [DYB98, CDSHD13]. This relies on accurate proxy geometry (usually point clouds or meshed), commonly enhanced via *Multi-View Stereo (MVS)* [SZPF16, GSC*07]. In real world datasets however, these techniques can present camera miscalibrations and erroneous geometry [SK00]. For image-based rendering, this can lead to warping artifacts, especially near object boundaries, or can cause blurring of details. Recently, pipelines enhanced by neural rendering [TTM*22] provided powerful tools to lessen these artifacts.

2.2. Neural Rendering and Scene Representations

In the last years, multiple variants of deep learning for novel view synthesis were introduced. Within proxy-based pipelines, several works have replaced the blending operation by a deep neural networks [RK21, HPP*18, RK20, FRF*23a] or learned textures [TZN19] during the warping stage. Other approaches use multi-plane images [MSOC*19, STB*19, TS20, ZTF*18] or estimate a warping fields [FNPS16, GKSL16, ZTS*16] to avoid the need of an scene specific proxy geometry.

This led the way for volumetric scene representations [PZ17] enhanced with deep learning [SMB*20, STH*19] and rendered via ray-marching. *Neural Radiance Fields (NeRFs)* [MST*21] furthermore showed that compressing a full 3D scene into a Multilayer Perceptron (MLP) achieve great results in this regard. This representation however is challenging in its own right, which follow-up works improve upon: long training times [CXZ*21, CBLPM21, MESK22, TMW*21, TRS22], many well distributed input views [CBLPM21, YYTK21, KD23] and rendering times [MESK22, BMT*21, NSP*21]. Improvements in quality [BMV*23, MBRS*21] allow NeRFs to surpass visual quality of many proxy-based approaches, however render times are still challenging, e.g. *MipNeRF-360* [BMV*22] ranging in the order of seconds per image and training needing dozens of hours.

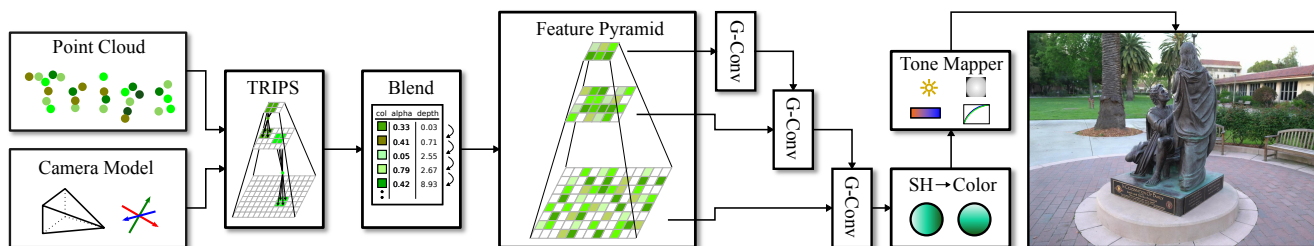


Figure 2: Our pipeline: TRIPS renders and blends a point cloud trilinearly as $2 \times 2 \times 2$ splats into multi-layered feature maps with the results being passed through our small neural network, containing only a single gated convolution per layer. Following, an optional spherical harmonics module and tone-mapper is used to produce the final image. This pipeline is completely differentiable, so that point descriptors (colors) and positions, as well as camera parameters are optimized via gradient descent.

Lately, discretizing parts of the scene space [YLT*21, HSM*21] or even replacing parts of it via voxel grids [FKYT*22], octrees [RWL*22] or tensor factorization [CXG*22] shrink computational costs as MLPs can be smaller or even removed. In this area, *InstantNGP* [MESK22] made waves as it uses hash-grids and a highly optimized MLP implementation for faster rendering and training speeds while retaining many qualitative advantages of NeRFs.

For the scope of real-time radiance field rendering however, Kerbl and Kopanas et al. [KKLD23] argue that ray-marching as a rendering concept is challenging on current GPU hardware.

2.3. Real-Time Rendering for Radiance Fields via Points

In the domain of real-time radiance field rendering, point clouds as an explicit proxy representation remain a great option. Point clouds are easily captured via LiDAR-based mapping [LXG22], RGB-D cameras with fusion techniques [DNZ*17, WSMG*16, KLL*13] and SfM/MVS techniques [SZPF16]. They represent a unstructured set of samples in space, with varying distances to neighbors, but true to the originally captured data. Rendering these can be very fast [SKW21, SKW22, SKW19], and augmenting points with neural descriptors [RFS22, ASK*20, RALB22, FRF*23b, HKT*23] or optimized attributes [KPLD21, KLR*22] provide high quality renderings using differentiable point renderers [WGSJ20, YSW*19] or neural ray-based renderers [XXP*22, OLN*22, ACDS24]. However, discrete rasterization of points can cause aliasing [SKW22] or overdraw [RFS22] if many points are rendered to the same pixel.

Another problem shared by point rendering techniques is how to fill holes in the unstructured data. Two main approaches have evolved over the years [KB04]: splatting (in world-space) and screen-space hole filling.

In world-space hole-filling, points are represented as oriented discs, often termed "splats" or "surfels", with disc radii pre-computed based on point cloud density. To reduce artifacts between neighboring splats, these discs can be rendered using Gaussian alpha-masks and combined with a normalizing blend function [AGP*04, PZVBG00, ZPVBG01]. Recent techniques optimize splat sizes [KKLD23, ZBRH22] or improve quality with neural networks [YCA*20]. For performance, overdraw poses a major issue as splats tend to overlap a lot. Thus, special care has to be

taken regarding the amount of splats drawn. 3D Gaussian Splatting [KKLD23] can be considered state of the art in this domain. They combine anisotropic Gaussians with a very fast tiled renderer and optimize splat sizes via gradient descent. However, limiting Gaussian numbers is necessary to avoid performance hits, which in turn can lead to over-blurring of small detailed elements.

The second direction involves screen-space hole-filling, where points, often rendered as tiny splats, are post-processed either through traditional methods [PGA11, MKC07, GD98] or using convolutional neural networks (CNNs) [ASK*20, MGK*19, SCCL20]. While these techniques bridge large point distances, their need for a large receptive field can result in artifacts or performance issues. A multi-resolution pyramid rendering approach mitigates this by assigning different network layers to varied resolutions [ASK*20, RALB22, HFF*23], albeit reintroducing overdraw issues at lower layers [RFS22]. Notably, ADOP [RFS22] excels in screen-space hole-filling, enabling the rendering of hundreds of millions of points for sharp object visualization [SKW22], but encounters challenges with temporal aliasing and substantial hole-filling.

Our approach aims to take the best of both worlds. Using TRIPS, we can render large splats by optimizing their size, but avoid high rasterization costs. This allows rendering enormous point clouds and detailed textures, while still being real-time capable without aliasing or temporal instability.

3. Method

Fig. 2 provides an overview of our rendering pipeline. The input data consists of images with camera parameters and a dense point cloud, which can be obtained through methods like multi-view stereo [SZPF16] or LiDAR sensing. To render a specific view, we project the neural color descriptors of each point into an image pyramid using the TRIPS technique (as detailed in Sec. 3.1) and blend them (Sec. 3.2). Subsequently, a compact neural reconstruction network (described in Sec. 3.3) integrates the layered representation, followed by the application of a spherical harmonics module (discussed in Sec. 3.4) and a tone mapper that transforms the resulting features into RGB colors.

Core to our method is the trilinear point renderer, which splats points bilinearly onto the screen space position as well as linearly to two resolution layers, determined by the projected point size.

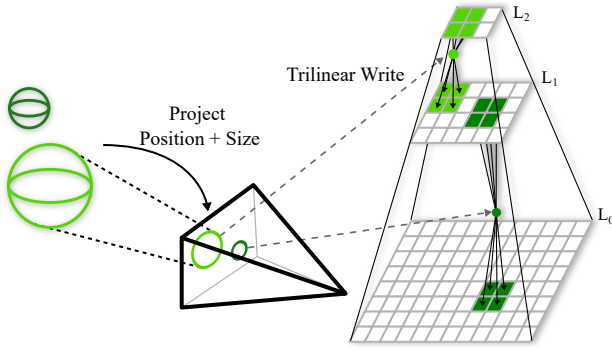


Figure 3: Trilinear Point Splatting: (left) all points and their respective size are projected into the target image. Based on this screen space size, each point is written to the correct layer of the image pyramid using a trilinear write (right). Large points are written to layers of lower resolution and therefore cover more space in the final image.

Our renderer uses similar nomenclature and is inspired by previous point-rasterizing approaches [SKW22, KPLD21, RFS22]. The neural image I is the output of the render function Φ

$$I = \Phi(C, R, t, x, E, s_w, \tau, \alpha), \quad (1)$$

where C are the camera intrinsics, (R, t) the extrinsic pose of the target view, x the position of the points, E the optional environment map, s_w the world space size of the points, τ the neural point descriptors and α the transparency for each point.

In contrast to other approaches, we do not use multiple render passes with progressively smaller resolutions, as this causes severe overdraw in the lower resolution layers. Instead, we compute the two layers which best match the point's projected size and render it only into these layers as 2×2 splat. By doing so, we mimic varying splat sizes, although effectively rendering only 2×2 -splats. The layers are then later merged in a small neural reconstruction network (Sec. 3.3) to the final image, resembling the decoder part of a U-Net.

3.1. Differentiable Trilinear Point Splatting

Using camera intrinsics C and pose (R, t) , we project each point position (x_w, y_w, z_w) to continuous (non-rounded) screen space coordinates (x, y, z) and each world-space point size s_w to screen space size s with the camera's focal length f :

$$s = \frac{f \cdot s_w}{z}. \quad (2)$$

Next, we render these points as a $2 \times 2 \times 2$ splats bilinearly and handle point size by splatting into two neighboring resolution layers L , as shown in Fig. 3. The resolution layers are selected to be the two closest in sizes to the projected size of the point with $L_{lower} = \lfloor \log(s) \rfloor$ and $L_{upper} = \lceil \log(s) \rceil$.

For each of the then selected eight pixels, we compute the contribution of the point to that pixel and augment its own transparency

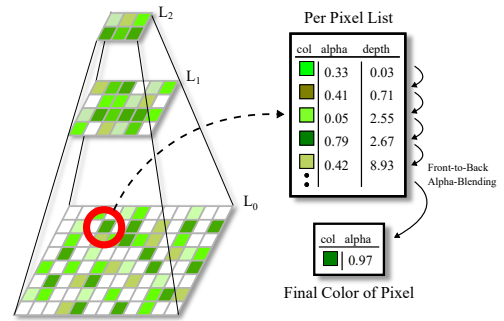


Figure 4: In each pixel of the image pyramid, a depth-sorted list of colors and alpha values is stored. The final color of each pixel is computed using front-to-back alpha blending on the sorted list.

value value with it. The final opacity value γ that is written to the image pyramid for pixel (x_i, y_i, s_i) with $s_i = 2^{L_i}$ is

$$\gamma = \beta \cdot \mathfrak{t} \cdot \alpha, \quad (3)$$

$$\beta = (1 - |x - x_i|) \cdot (1 - |y - y_i|) \quad (4)$$

$$\mathfrak{t} = \begin{cases} 1 - \frac{|s - s_i|}{2^{L_{upper}} - 2^{L_{lower}}} & s \geq 1 \\ \epsilon + (1 - \epsilon)s & s_i = 0 \wedge s < 1 \end{cases} \quad (5)$$

where β is the bilinear weight inside the image layer, \mathfrak{t} is the linear layer weight, and α the opacity value of the point. The layer weight \mathfrak{t} is a standard linear interpolation if the point size s is inside the image pyramid. The second case of Equ. (5) handles far away points that have a pixel size smaller than one. In order not to miss these, we always add them to the finest level 0. To avoid that their weight disappears, we ensure that their contribution is at least $\epsilon = 0.25$.

3.2. Multi Resolution Alpha Blending

Since each point is written to multiple pixels and multiple points can fall into the same pixel, we collect all fragments in per pixel lists Λ_{i, x_i, y_i} . These lists are sorted by depth and clamped to a maximum size of 16 elements. Eventually, the color C_Λ is computed using front-to-back alpha blending (Fig. 4):

$$C_\Lambda = \sum_{m=1}^{|\Lambda|} T_m \cdot \alpha_m \cdot c_m \quad (6)$$

$$T_m = \prod_{i=1}^{m-1} (1 - \alpha_i), \quad (7)$$

3.3. Neural Network

The result produced by our renderer consists of a feature image pyramid comprising n layers. These individual layers are finally consolidated into a single full-resolution image by a compact neural network, as depicted in Fig. 2. Our network architecture incorporates a single gated convolution [YLY*19] in each layer with a self-bypass connection and a feature size of 32. Additionally, we include a bilinear upsampling operation for all layers except the final

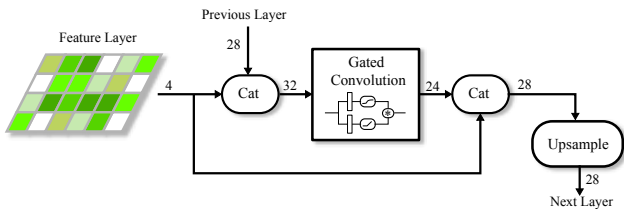


Figure 5: Our design of one gated convolution block that processes the features of the image pyramid with the number of channels passed through indicated at each step.

one, merging the output with the subsequent level. This configuration is shown in Fig. 5 and resembles an efficient decoder network, due to its restrained number of features, pixels, and convolutional operations.

Unlike well-established hole-filling neural networks [ASK*20, RFS22, RALB22], our approach demands a significantly smaller and more efficient network. This reduced network size stems from the fact that our renderer is adept at filling gaps autonomously and generates smooth output through trilinearly splatting points. Consequently, the network’s primary task is to learn minimal hole-filling and outlier removal, allowing it to concentrate its efforts on high-quality texture reconstruction.

3.4. Spherical Harmonics Module and Tone Mapping

To model view dependent effects and camera-specific capturing parameters (like exposure time), we optionally interpret the network output as spherical harmonics (SH) coefficients, convert them to RGB colors, and finally pass the result to a physically-based tone mapper. This allows the system to make use of explicit view directions. The SH-module makes use of spherical harmonics with degree 2, which corresponds to 27 input coefficients (9 coefficients per color channel). These coefficients are the output of the last convolution of our network. The tone mapper follows the work of Rückert et al. [RFS22], which models exposure time, white balance, sensor response, and vignetting.

3.5. Optimization Strategy

Before novel views can be synthesized, the rendering pipeline is optimized to reproduce the input photographs. This optimization includes point position, size, and features, as well as the camera model and poses, neural network weights, and tone mapper parameters. We train for 600 epochs, which, depending on scene size, requires 2-4 hours to converge.

As training criterion, we use the VGG-loss [JAF16] which has been shown to provide high-quality results [RFS22]. The VGG network, however, tends to be slow to evaluate, thus increasing training times significantly compared to MSE loss. Therefore, we use a combination of MSE and SSIM [KKLD23] in the first 50 epochs when the advantages of VGG are still negligible. This speeds up training time by about 5% percent.

Similar to Kerbl and Kopanas [KKLD23], we use a "warm-up" period of 20 epochs, during which we train with half image resolutions. Afterwards we randomly zoom in and out each epoch, so that all convolutions (whose weights are not shared) are trained to contribute to the final result.

3.6. Implementation Details

Our implementation uses *torch* as auto-differentiable backend, however the trilinear renderer is implemented in custom CUDA kernels, as they commonly provide better performance [KKLD23, RFS22]. Fast spherical harmonics encodings are provided by *tiny-cuda-nn* [Mül21].

The renderer is implemented in three stages: collecting, splatting and accumulation, albeit diverging from other state-of-the-art multi-layer blending strategies, this turned out to work best in our scenario [FHSS18, LZ21, VVP20]. We first project each point (x_w, y_w, z_w) to the desired view and collect each point’s (x, y, z) as well as point size s in a buffer, and also count how many elements are mapped to each pixel. This counting is then used for an offset scan to index into one continuous arrays for all layers. The following splatting pass duplicates each point and stores a pair of (z, i) (with i an index to the stored information) in each pixels’ list.

Following, a combined sorting and accumulation pass is done. Regarding performance, this part is critical, as such we opt to only use the front most 16 elements from each (sorted) list, a common practice when blending points [LZ21]. We could not identify any loss of quality caused by this approximation, as the blending contribution of later points is very low. This limitation allows us to use GPU-friendly sorting, as we repeat warp-local (32 threads) and shuffle-based bitonic sorts, always replacing the latter 16 elements with new unsorted ones, until the lists are empty. For the backwards pass, the sorted per-pixel lists are stored, allowing fast backpropagation. The front-to-back alpha blending (see Sec. 3.2) is done in the same pass as the sorting pass, because all relevant elements are already in registers.

In contrast to Kerbl and Kopanas et al. [KKLD23], we use this per-pixel sorting, which proved to be faster for us than global sorting. This is mostly due to the higher amount and smaller sizes of points in our approach.

For scenes with a large deviation in point density, we found that occlusion may not be correctly evaluated by the neural network in edge cases. Therefore, we include points from coarser layers during blending (in the usual way), of which the additional cost is very small ($< 0.5\text{ms}$).

Point sizes are initialized with the average distance to the four nearest neighbor, which is then efficiently optimized during training (see Fig. 6).

4. Evaluation

Next, we compare our approach with prior arts as well as showcase the effectiveness of our design decisions in ablation studies.

Table 1: Results on the Tanks&Temples and MipNeRF-360 datasets, as well as BOAT and OFFICE. See also Fig. 7 for visual comparisons.

Method	Tanks&Temples			MipNeRF-360			BOAT			OFFICE		
	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑
InstantNGP	0.475	21.74	0.692	0.374	25.94	0.697	0.598	15.34	0.455	0.544	13.45	0.801
Mip-NeRF 360	0.340	24.61	0.789	0.286	28.23	0.796	0.680	12.20	0.357	0.526	15.22	0.832
Gaussian Spl.	0.300	24.63	0.818	0.278	26.94	0.792	0.544	15.30	0.470	0.371	18.77	0.878
ADOP	0.229	23.78	0.802	0.285	23.26	0.707	0.301	20.49	0.650	0.279	21.47	0.899
Ours	0.213	24.64	0.808	0.233	25.94	0.772	0.301	20.38	0.633	0.271	21.36	0.887

4.1. Setup and Datasets

We have evaluated our approach on several scenes from the Tanks&Temples [KPZK17] and the MipNeRF-360 [BMV*22] datasets. Additionally, we use the BOAT and OFFICE scene from Rückert et al. [RFS22] to evaluate robustness towards difficult input conditions. The former contains outdoor auto-exposed images while the later is a office floor with multiple distinct room and a large LiDAR point cloud, but sparsely placed cameras.

From Tanks&Temples, we use the *intermediate* set containing eight scenes: TRAIN, PLAYGROUND, M60, LIGHTHOUSE, FAMILY, FRANCIS, HORSE and PANTHER. These scenes are outdoor scenes captured under varying lighting conditions but with good spatial coverage and can be seen as a good baseline for robustness. The MipNeRF-360 dataset [BMV*22] consists of 5 outdoor and 4 indoor scenes. This dataset was captured with controlled setups and has capture positions well suited for volumetric rendering with a hemispherical setup [KD23]. We use half resolution for images of this dataset, resulting in resolutions of around 2500×1600 px for outdoor and 1550×1030 px for indoor scenes. For results with the resolutions used in related works (quarter resolution for outdoor and half resolution for indoor), take a look at the supplemental document, Tabs. 1-3.

Point clouds of all scenes were acquired via COLMAP’s MVS [SZPF16], except OFFICE which was captured by LiDAR.

For the quantitative evaluation we use the LPIPS_{VGG} [ZIE*18], PSNR, and SSIM metrics. We note however, that neither of these metrics always reflect visual impression. Some approaches are trained with MSE-loss or SSIM and therefore naturally perform better in PSNR and SSIM. Our approach, on the other hand, is trained with VGG-loss and thus usually shows better scores on LPIPS. For a fair comparison, we recommend to look at all metrics and closely inspect the provided image and video comparison.

In all experiments, we leave every 8th view out for testing. This is the same train/test split as used in current related work [BMV*22, KKLD23].

4.2. Quality Comparison

In Tab. 1 and Fig. 7, we compare our approach to Instant-NGP [MESK22], MipNeRF-360 [BMV*22], 3D Gaussian Splatting [KKLD23] and ADOP [RFS22]. The latter two are the closest-related point-based radiance field rendering approaches.

On the Tanks&Temples dataset, our approach achieves in average the best LPIPS score with an improvement of 20% over the

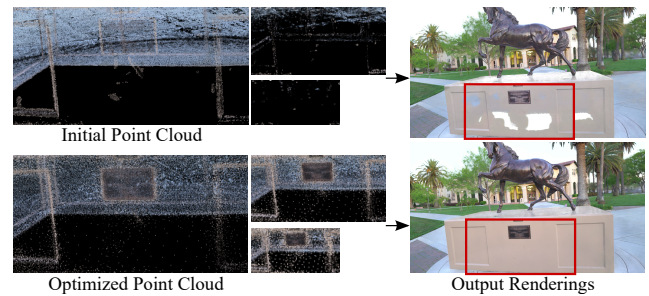


Figure 6: The initial COLMAP reconstruction lacks point on the pedestal of the statue (top left). Our approach distributes the few present points and increases their sizes (bottom left) thus rendering them also in lower layers (middle). Thus our pipeline can avoid distracting holes (right).

second best. In PSNR and SSIM the score is on par with state-of-the-art. On the MipNeRF-360 dataset, we obtain again the best LPIPS score, however the volumetric methods and Gaussian Splatting show an improved PSNR and SSIM. The difference can be inspected in Fig. 7. For example, in row 3, the TRIPS rendering provides better sharpness with more details, but the MipNeRF-360 and Gaussian output is overall cleaner with less noise. On the difficult BOAT and OFFICE scenes, we can show that our rendering pipeline, is robust to extreme input conditions.

Individual scores per scene can be seen in the supplemental document Tabs. 5-10.

4.3. Ablation Studies

In this section, first we show the effect of our design choices.

4.3.1. Point-Size Optimization

With our trilinear splatting technique, point sizes can be optimized to fill large holes in the scene. We show this capability in Fig. 6, where the initial point cloud exhibits a large hole in the pedestal of the horse producing artifacts in rendering (top row). To combat this, our pipeline efficiently moves and enlarges the points to fill the hole (bottom row), thus providing great render quality.

4.3.2. Point Position Optimization

To test the efficiency of our trilinear point position optimization compared to the (cheaper) approximate gradients from ADOP, we added random noise (of 0.01) to the positions of all points after

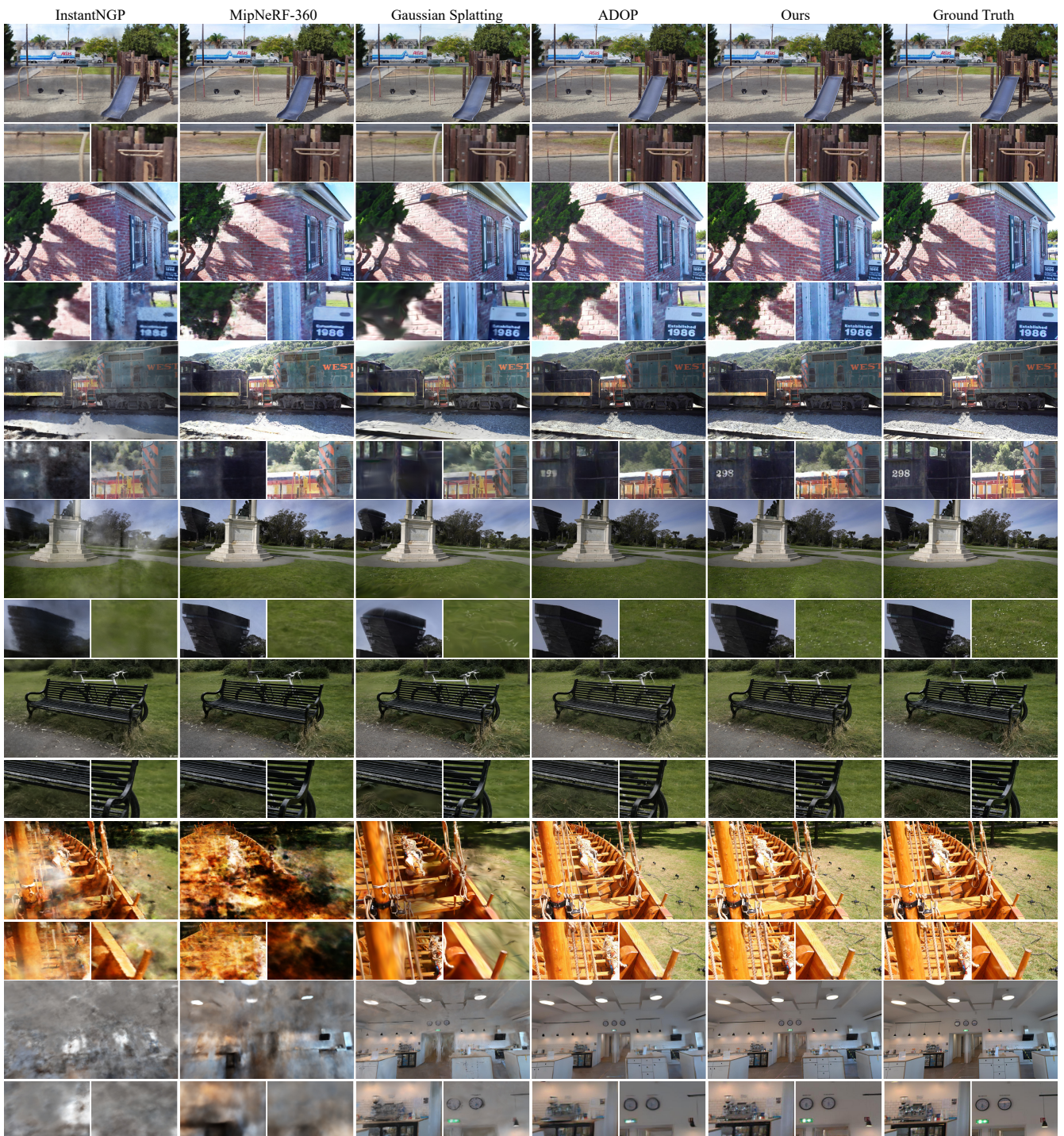


Figure 7: Visual comparisons.

Table 2: View dependency on different scenes. On scenes with strong view dependency (GARDEN), adding view dependant configurations, either via our SH network module (SH-net) or optimized per point (SH-point) increases quality, however the per-point point setup severely impacts performance. Our module gives a balanced trade off, which also avoids over-fitting on less view-dependent scenes (PLAYGROUND).

View-dep	PLAYGROUND (12.5M Points)				HORSE (1.8M Points)				GARDEN (8.2M Points)			
	LPIPS ↓	PSNR ↑	SSIM ↑	Time ↓	LPIPS ↓	PSNR ↑	SSIM ↑	Time ↓	LPIPS ↓	PSNR ↑	SSIM ↑	Time ↓
none	0.225	24.85	0.720	11.1ms	0.202	22.73	0.822	7.7ms	0.219	24.82	0.756	17.9ms
SH-net	0.225	24.88	0.724	13.3ms	0.203	22.81	0.825	8.9ms	0.222	24.46	0.752	18.5ms
SH-point	0.236	24.32	0.702	27.4ms	0.200	22.89	0.829	10.2ms	0.213	25.15	0.756	27.3ms

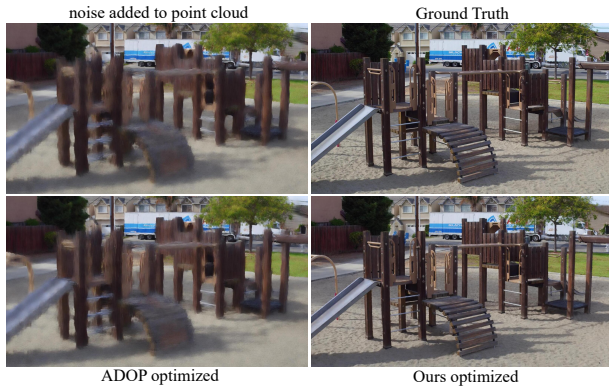


Figure 8: We added noise to the converged point clouds of ADOP and ours, then restarted optimization for positions only. Ours is able to converge back to the correct result, ADOP fails at that.

Table 3: Number of resolution layers used (HORSE scene).

#Layers	LPIPS ↓	PSNR ↑	SSIM ↑	Time ↓
3	0.216	21.24	0.818	7.10ms
4	0.201	22.40	0.826	7.35ms
5	0.197	22.76	0.826	7.42ms
6	0.196	23.06	0.829	7.50ms
7	0.195	23.10	0.826	7.54ms
8	0.192	23.34	0.828	7.61ms

training, then optimize only point positions for 100 epochs. The result can be seen in Fig. 8. Our pipeline is able to reconstruct the correct rendering, while ADOP’s result barely improves.

4.3.3. Number of Render Layers

Due to our trilinear point rendering algorithm, increasing the number of pyramid layers has almost no negative impact on render time. As seen in Tab. 3, having 8 layers improves quality, especially with PSNR. For reference, other approaches make use of 4 [RFS22] or 5 [ASK*20] layers and describe significant performance impacts when increasing the number of layers [RFS22].

4.3.4. View Dependency

After the neural network, optionally we use a spherical harmonics module to model view depended artifacts of the scene. This improves the rendering quality for some scenes (GARDEN), while for others it makes little to no difference (see Tab. 2). Applying the spherical harmonics before the network achieves roughly the same

Table 4: Features per point on the PLAYGROUND scene.

# Features	LPIPS ↓	PSNR ↑	SSIM ↑	Time ↓
4	0.225	24.85	0.720	11.1ms
6	0.231	24.61	0.701	11.7ms
8	0.223	25.04	0.727	12.2ms

Table 5: Network configuration compared (PLAYGROUND scene).

Network	LPIPS ↓	PSNR ↑	SSIM ↑	Time ↓
ADOP-net	0.236	24.74	0.713	10.7ms
ours	0.225	24.85	0.720	4.5ms
ours+SH	0.225	24.88	0.724	5.6ms
ours+SH _{L2}	0.248	24.34	0.684	2.2ms

quality, but also reduces efficiency due to additional memory overhead. On scenes without reflective materials, skipping the spherical harmonics module is thus possible.

4.3.5. Feature Vector Dimensions

Our pipeline uses by default four feature descriptors per point. More features only marginally increase the quality, while requiring significantly more memory and slightly increasing rendering time, as shown in Tab. 4.

4.3.6. Networks

In our pipeline, we use a small decoder network made out of gated convolutions, presented in Sec. 3.3. ADOP [RFS22] on the other hand uses a four layer U-net with double convolutions for encoder and decoder (thus around 6 times more parameter). As seen in Tab. 5, in our pipeline our networks provide similar quality to ADOP’s full network, while being much faster in inference. With spherical harmonics, inference times slightly increase, but the system is now able to model view dependency. Adding the SH-module to the second finest layer (ours+SH_{L2}) instead of the finest (ours+SH) of the network improves efficiency but weakens results.

4.3.7. Time Scaling on Number of Points

As seen in Tab. 6, TRIPS is very efficient in rendering large amounts of points. Even for our largest scene with more than 70M

Table 6: Efficiency of our approach regarding point cloud sizes.

Scene	HORSE	GARDEN	PLAYGR.	BOAT	OFFICE
#Points	1.8M	7.8M	12.5M	53.0M	72.5M
Time	2.5ms	5.9ms	6.2ms	13.1ms	15.0ms

Table 7: Training and render times on the GARDEN (images resolution: 2594×1681) and PLAYGROUND scene (1920×1080).

Method	Train	Render(GARDEN)	Render(PLAYGR.)
InstantNGP	0.25h	131ms	172ms
Mip-NeRF360	36h	38000ms	18000ms
ADOP	8h	30.3ms	14.5ms
Gaussian Spl.	0.75h	11.5ms	8.6ms
Ours	4h	16.4ms	11.1ms

Table 8: Breakdown of the frame time for the PLAYGROUND scene. Our method's "Rasterize" consists of: counting and memory allocation with 1.9ms, splatting with 2.6ms and combined sorting and blending with 1.7ms.

Method	#Points	Rasterize	Network	Tonemap	In Total
ADOP	12M	3.1ms	11.0ms	0.4ms	14.5ms
Gauss. Spl.	2M	8.6ms			8.6ms
Gauss. Spl.	8M	11.4ms			11.4ms
Ours	12M	6.2ms	4.5ms	0.4ms	11.1ms

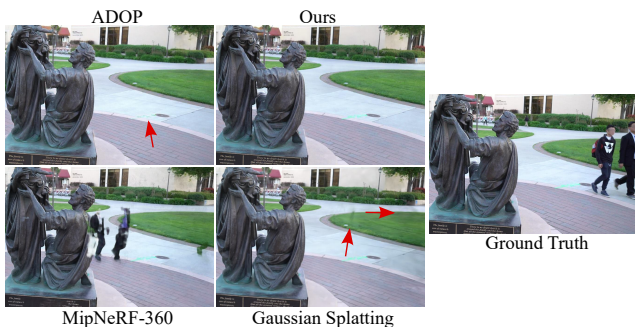
points, the pipeline remains real-time capable with only 15ms required for rasterization.

4.4. Rendering Efficiency

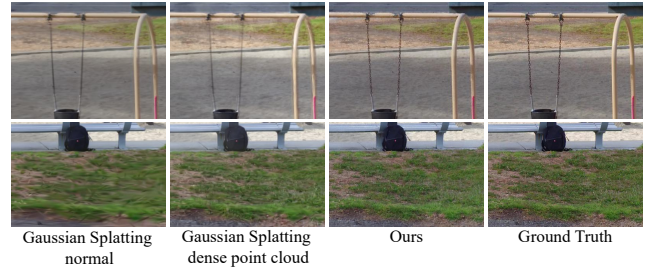
In Tab. 7, we evaluate training and rendering time for all examined methods. Our method trains for around 2-4h per scene on an Nvidia A100 and renders a novel view in around 11ms on an RTX4090. A finer breakdown of the steps involved can be found in Tab. 8.

4.5. Outlier Robustness

As seen in Fig. 9, our approach is robust to outlier measurements, for example, people walking through the scene. Especially volumetric approach like MipNeRF-360, suffer from severe artifacts in this case, due to strong view-dependant over-fitting capability.

**Figure 9:** Comparison of outlier robustness on the FAMILY scene. Only our methods is able to remove floating artifacts while still retaining full color precision on the sidewalk.**Table 9:** Performance of the methods on the PLAYGROUND scene. Gaussian (dense) starts with COLMAP's dense reconstruction of 12M points and prunes them to 8M, Gaussian (sparse) is the original sparse setup and has about 2M points. Also see Fig. 10.

Method	LPIPS↓	PSNR↑	SSIM↑	Time ↓
Ours	0.229	25.12	0.746	11.1ms
ADOP	0.233	24.86	0.753	14.5ms
Gaussian (dense)	0.283	24.06	0.773	11.4ms
Gaussian (sparse)	0.322	24.61	0.776	8.6ms

**Figure 10:** Visual results of Gaussian splatting with COLMAP's dense point cloud as input compared its normal setup as well as ours, which provides the sharpest results (PLAYGROUND scene).

4.6. Comparison to Prior Work with Number of Points

We have seen in previous experiments that Gaussian Splatting [KKLD23] has blurrier results compared to TRIPS, which can be confirmed by their weak LPIPS scores. However, they start with fewer point primitives (the SfM reconstruction) and thus are limited in the amount of detail to display. To this end, we conducted an experiment, where the Gaussian Splatting pipeline is provided with the dense point cloud (providing the same input as for our pipeline). Gaussian splatting has a pruning mechanism to remove unwanted Gaussian, thus after their full training, from the initial 12.5M points only around 8M survived.

The results of this experiment are presented in Tab. 9. It can be seen that LPIPS improves with more Gaussians (however PSNR declines) as fine details can be reconstructed better. The qualitative comparison paints the same picture (see Fig. 10), where the quality of the grass improves drastically, however finer details such as the chains still can only be reconstructed by us. Overall the technique cannot reach the quality and scores of TRIPS, as we can keep more points to render efficiently as well as use neural descriptors to encode more detailed information.

Furthermore, our approach performs more efficiently in scenarios with large point clouds. In the dense setup, TRIPS outperforms Gaussian Splatting, as the resolution-dependant computation cost of our neural network (4.5ms at 1920×1080) catches up with our more efficient point rasterizer (see Tab. 8).

5. Limitations

In the preceding section, we have demonstrated TRIPS' effectiveness on commonly encountered real-world datasets. Nonetheless, we have also identified potential limitations. One such limitation



Figure 11: Limitation: Holefilling close to the camera exhibits fuzzy edges and shine-through.

arises from the prerequisite to have an initial dense reconstruction (in contrast to Gaussian Splatting), which may not be practical in certain scenarios.

Additionally, our lack of an anisotropic splat formulation can create problems: When our method is tasked with strong holefilling of elongated, slender object (such as poles), noisy artifacts surrounding their silhouettes can be observed. An example of this is depicted in Fig. 11. In such instances, the slightly blurred edge characteristic of Gaussian Splatting is often preferred.

Furthermore, even though the temporal consistency compared to previous point rendering approaches [ASK*20, RFS22] has been drastically improved, slight flickering can still occur in areas with too many or too little points.

Our trilinear point splatting splits up points into distinct layers and as such loses depth information. Theoretically, during recombination this could create holes in solid geometry. In practice, we could not find instances of this happening except in extreme zoom-ins far outside the training data. We believe that the per-point descriptors, the point inclusion in coarse layers, and the network-based recombination are capable to combat this issue, as reflected in the rendering quality.

6. Conclusion

In this paper, we presented TRIPS, a robust real-time point-based radiance field rendering pipeline. TRIPS employs an efficient strategy of rasterizing points into a screen-space image pyramid, allowing the efficient rendering of large points and is completely differentiable, thus allowing automatic optimization of point sizes and positions. This technique enables the rendering of highly detailed scenes and the filling of large gaps, all while maintaining a real-time frame rate on commonly available hardware.

We highlight that TRIPS achieves high rendering quality, even in challenging scenarios like scenes with intricate geometry, large-scale environments, and auto-exposed footage. Moreover, due to the smooth point rendering approach, a comparably simple neural reconstruction network is sufficient, resulting in real-time rendering performance.

An open source implementation is available under:

<https://github.com/lfranke/TRIPS>

Acknowledgements

We thank the anonymous reviewers for their valuable feedback. We thank Matthias Innmann, Stefan Romberg, Michael Gerstmayer and Tim Habigt for the fruitful discussions as well as NavVis GmbH for providing the Office dataset.

Linus Franke was supported by the Bayerische Forschungstiftung (Bavarian Research Foundation) AZ-1422-20. The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project *b162dc*. NHR funding is provided by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by the German Research Foundation (DFG) – 440719683. Open Access funding enabled and organized by Projekt DEAL.

References

- [ACDS24] ABOU-CHAKRA J., DAYOUB F., SÜNDERHAUF N.: Particleclerf: A particle-based encoding for online neural radiance fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2024), pp. 5975–5984. 3
- [AGP*04] ALEXA M., GROSS M., PAULY M., PFISTER H., STAMMINGER M., ZWICKER M.: Point-based computer graphics. In *ACM SIGGRAPH 2004 Course Notes*. 2004, pp. 7–es. 3
- [ASK*20] ALIEV K.-A., SEVASTOPOLSKY A., KOLOS M., ULYANOV D., LEMPITSKY V.: Neural point-based graphics. In *European Conference on Computer Vision* (2020), Springer, pp. 696–712. 2, 3, 5, 8, 10
- [BMT*21] 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. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2021), pp. 5855–5864. 2
- [BMV*22] 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* (2022), pp. 5470–5479. 2, 6
- [BMV*23] BARRON J. T., MILDENHALL B., VERBIN D., SRINIVASAN P. P., HEDMAN P.: Zip-nerf: Anti-aliased grid-based neural radiance fields. *arXiv preprint arXiv:2304.06706* (2023). 2
- [CBLPM21] CHIBANE J., BANSAL A., LAZOVA V., PONS-MOLL G.: Stereo radiance fields (srf): Learning view synthesis for sparse views of novel scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 7911–7920. 2
- [CDSHD13] CHAURASIA G., DUCHENE S., SORKINE-HORNUNG O., DRETTAKIS G.: Depth synthesis and local warps for plausible image-based navigation. *ACM Transactions on Graphics (TOG)* 32, 3 (2013), 1–12. 2
- [CXG*22] CHEN A., XU Z., GEIGER A., YU J., SU H.: Tensorf: Tensorial radiance fields. In *Computer Vision – ECCV 2022* (Cham, 2022), Avidan S., Brostow G., Cissé M., Farinella G. M., Hassner T., (Eds.), Springer Nature Switzerland, pp. 333–350. 3
- [CXZ*21] CHEN A., XU Z., ZHAO F., ZHANG X., XIANG F., YU J., SU H.: Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 14124–14133. 2
- [DNZ*17] DAI A., NIESSNER M., ZOLLHÖFER M., IZADI S., THEOBALT C.: Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1. 3

- [DYB98] DEBEVEC P., YU Y., BOSHOKOV G.: Efficient view-dependent ibr with projective texture-mapping. In *EG Rendering Workshop* (1998), vol. 4. 2
- [FHSS18] FRANKE L., HOFMANN N., STAMMINGER M., SELGRAD K.: Multi-layer depth of field rendering with tiled splatting. *Proceedings of the ACM on Computer Graphics and Interactive Techniques 1*, 1 (2018), 1–17. 5
- [FKYT*22] FRIDOVICH-KEIL S., YU A., TANCIK M., CHEN Q., RECHT B., KANAZAWA A.: Plenoxels: Radiance fields without neural networks. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 5491–5500. 3
- [FNPS16] FLYNN J., NEULANDER I., PHILBIN J., SNAVELY N.: Deepstereo: Learning to predict new views from the world’s imagery. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 5515–5524. 2
- [FRF*23a] FINK L., RÜCKERT D., FRANKE L., KEINERT J., STAMMINGER M.: Liveness: Neural view synthesis on live rgb-d streams. In *SIGGRAPH Asia Conference Papers* (New York, NY, USA, Dec. 2023), Association for Computing Machinery. 2
- [FRF*23b] FRANKE L., RÜCKERT D., FINK L., INNMANN M., STAMMINGER M.: Vet: Visual error tomography for point cloud completion and high-quality neural rendering. In *SIGGRAPH Asia Conference Papers* (New York, NY, USA, Dec. 2023), Association for Computing Machinery. 3
- [GD98] GROSSMAN J. P., DALLY W. J.: Point sample rendering. In *Eurographics Workshop on Rendering Techniques* (1998), Springer, pp. 181–192. 3
- [GKSL16] GANIN Y., KONONENKO D., SUNGATULLINA D., LEMPITSKY V.: Deepwarp: Photorealistic image resynthesis for gaze manipulation. In *European conference on computer vision* (2016), Springer, pp. 311–326. 2
- [GSC*07] GOESELE M., SNAVELY N., CURLESS B., HOPPE H., SEITZ S. M.: Multi-view stereo for community photo collections. In *2007 IEEE 11th International Conference on Computer Vision* (2007), IEEE, pp. 1–8. 2
- [HFF*23] HARRER M., FRANKE L., FINK L., STAMMINGER M., WEYRICH T.: Inovis: Instant novel-view synthesis. In *SIGGRAPH Asia Conference Papers* (New York, NY, USA, Dec. 2023), Association for Computing Machinery. 3
- [HKT*23] HAHLBOHM F., KAPPEL M., TAUSCHER J.-P., EISEMANN M., MAGNOR M.: Plenopticpoints: Rasterizing neural feature points for high-quality novel view synthesis. In *Proc. Vision, Modeling and Visualization (VMV)* (2023), Eurographics. 3
- [HPP*18] HEDMAN P., PHILIP J., PRICE T., FRAHM J.-M., DRETTAKIS G., BROSTOW G.: Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–15. 2
- [HSM*21] HEDMAN P., SRINIVASAN P. P., MILDENHALL B., BARRON J. T., DEBEVEC P.: Baking neural radiance fields for real-time view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 5875–5884. 3
- [JAF16] JOHNSON J., ALAHI A., FEI-FEI L.: Perceptual losses for real-time style transfer and super-resolution. *CoRR abs/1603.08155* (2016). 5
- [KB04] KOBBELT L., BOTSCH M.: A survey of point-based techniques in computer graphics. *Computers & Graphics* 28, 6 (2004), 801–814. 3
- [KD23] KOPANAS G., DRETTAKIS G.: Improving NeRF Quality by Progressive Camera Placement for Free-Viewpoint Navigation. In *Vision, Modeling, and Visualization* (2023), Guthe M., Grosch T., (Eds.), The Eurographics Association. 2, 6
- [KKLD23] KERBL B., KOPANAS G., LEIMKÜHLER T., DRETTAKIS G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* 42, 4 (2023). 1, 2, 3, 5, 6, 9
- [KLL*13] KELLER M., LEFLOCH D., LAMBERS M., IZADI S., WEYRICH T., KOLB A.: Real-time 3D reconstruction in dynamic scenes using point-based fusion. In *Proc. of Joint 3DIM/3DPVT Conference (3DV)* (June 2013), pp. 1–8. Selected for oral presentation. 3
- [KLR*22] KOPANAS G., LEIMKÜHLER T., RAINER G., JAMBON C., DRETTAKIS G.: Neural point catacaustics for novel-view synthesis of reflections. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–15. 2, 3
- [KPLD21] KOPANAS G., PHILIP J., LEIMKÜHLER T., DRETTAKIS G.: Point-based neural rendering with per-view optimization. In *Computer Graphics Forum* (2021), vol. 40, Wiley Online Library, pp. 29–43. 2, 3, 4
- [KPZK17] KNPITSCH A., PARK J., ZHOU Q.-Y., KOLTUN V.: Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics* 36, 4 (2017). 6
- [LXG22] LIAO Y., XIE J., GEIGER A.: Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 3 (2022), 3292–3310. 3
- [LZ21] LASSNER C., ZOLLHÖFER M.: Pulsar: Efficient sphere-based neural rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2021), pp. 1440–1449. 5
- [MBRS*21] 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* (2021), pp. 7210–7219. 2
- [MESK22] MÜLLER T., EVANS A., SCHIED C., KELLER A.: Instant neural graphics primitives with a multiresolution hash encoding. *arXiv preprint arXiv:2201.05989* (2022). 2, 3, 6
- [MGK*19] MESHRY M., GOLDMAN D. B., KHAMIS S., HOPPE H., PANDEY R., SNAVELY N., MARTIN-BRUALLA R.: Neural rerendering in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 6878–6887. 3
- [MKC07] MARROQUIM R., KRAUS M., CAVALCANTI P. R.: Efficient point-based rendering using image reconstruction. In *PBG@ Eurographics* (2007), pp. 101–108. 3
- [MSOC*19] MILDENHALL B., SRINIVASAN P. P., ORTIZ-CAYON R., KALANTARI N. K., RAMAMOORTHY R., NG R., KAR A.: Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14. 2
- [MST*21] MILDENHALL B., SRINIVASAN P. P., TANCIK M., BARRON J. T., RAMAMOORTHY R., NG R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* 65, 1 (2021), 99–106. 2
- [Mül21] MÜLLER T.: tiny-cuda-nn, 4 2021. URL: <https://github.com/NVlabs/tiny-cuda-nn>. 5
- [NSP*21] NEFF T., STADLBAUER P., PARGER M., KURZ A., MUELLER J. H., CHAITANYA C. R. A., KAPLAYAN A., STEINBERGER M.: Donerf: Towards real-time rendering of compact neural radiance fields using depth oracle networks. In *Computer Graphics Forum* (2021), vol. 40, Wiley Online Library, pp. 45–59. 2
- [OLN*22] OST J., LARADJI I., NEWELL A., BAHAT Y., HEIDE F.: Neural point light fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 18419–18429. 3
- [PGA11] PINTUS R., GOBBETTI E., AGUS M.: Real-time rendering of massive unstructured raw point clouds using screen-space operators. In *Proceedings of the 12th International conference on Virtual Reality, Archaeology and Cultural Heritage* (2011), pp. 105–112. 3
- [PZ17] PENNER E., ZHANG L.: Soft 3d reconstruction for view synthesis. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 1–11. 2

- [PZVBG00] PFISTER H., ZWICKER M., VAN BAAR J., GROSS M.: Surfels: Surface elements as rendering primitives. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), pp. 335–342. 3
- [RALB22] RAKHIMOV R., ARDELEAN A.-T., LEMPITSKY V., BURNAEV E.: NPBG++: accelerating neural point-based graphics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (June 2022), pp. 15969–15979. 3, 5
- [RFS22] RÜCKERT D., FRANKE L., STAMMINGER M.: Adop: Approximate differentiable one-pixel point rendering. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–14. 1, 2, 3, 4, 5, 6, 8, 10
- [RK20] RIEGLER G., KOLTUN V.: Free view synthesis. In *European Conference on Computer Vision* (2020), Springer, pp. 623–640. 2
- [RK21] RIEGLER G., KOLTUN V.: Stable view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 12216–12225. 2
- [RWL*22] RÜCKERT D., WANG Y., LI R., IDOUGH R., HEIDRICH W.: Neat: Neural adaptive tomography. *ACM Trans. Graph.* 41, 4 (2022). 3
- [SCCL20] SONG Z., CHEN W., CAMPBELL D., LI H.: Deep novel view synthesis from colored 3d point clouds. In *European Conference on Computer Vision* (2020), Springer, pp. 1–17. 3
- [SF16] SCHONBERGER J. L., FRAHM J.-M.: Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 4104–4113. 2
- [SK00] SHUM H., KANG S. B.: Review of image-based rendering techniques. In *Visual Communications and Image Processing 2000* (2000), vol. 4067, SPIE, pp. 2–13. 2
- [SKW19] SCHÜTZ M., KRÖSL K., WIMMER M.: Real-time continuous level of detail rendering of point clouds. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2019), IEEE, pp. 103–110. 3
- [SKW21] SCHÜTZ M., KERBL B., WIMMER M.: Rendering point clouds with compute shaders and vertex order optimization. In *Computer Graphics Forum* (2021), vol. 40, Wiley Online Library, pp. 115–126. 3
- [SKW22] SCHÜTZ M., KERBL B., WIMMER M.: Software rasterization of 2 billion points in real time. *arXiv preprint arXiv:2204.01287* (2022). 3, 4
- [SMB*20] SITZMANN V., MARTEL J., BERGMAN A., LINDELL D., WETZSTEIN G.: Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems* 33 (2020). 2
- [SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3d. In *ACM siggraph 2006 papers*. 2006, pp. 835–846. 2
- [STB*19] SRINIVASAN P. P., TUCKER R., BARRON J. T., RAMAMOORTHY R., NG R., SNAVELY N.: Pushing the boundaries of view extrapolation with multiplane images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 175–184. 2
- [STH*19] SITZMANN V., THIES J., HEIDE F., NIESSNER M., WETZSTEIN G., ZOLLHOFER M.: Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 2437–2446. 2
- [SZPF16] 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). 2, 3, 6
- [TMW*21] TANCİK M., MILDENHALL B., WANG T., SCHMIDT D., SRINIVASAN P. P., BARRON J. T., NG R.: Learned initializations for optimizing coordinate-based neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 2846–2855. 2
- [TRS22] TURKI H., RAMANAN D., SATYANARAYANAN M.: Meganerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 12922–12931. 2
- [TS20] TUCKER R., SNAVELY N.: Single-view view synthesis with multiplane images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 551–560. 2
- [TTM*22] TEWARI A., THIES J., MILDENHALL B., SRINIVASAN P., TRETSCHK E., YIFAN W., LASSNER C., SITZMANN V., MARTINBRUALLA R., LOMBARDI S., ET AL.: Advances in neural rendering. In *Computer Graphics Forum* (2022), vol. 41, Wiley Online Library, pp. 703–735. 2
- [TZN19] THIES J., ZOLLHÖFER M., NIESSNER M.: Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–12. 2
- [VVP20] VASILAKIS A.-A., VARDIS K., PAPAIOANNOU G.: A survey of multiframe rendering. In *Computer Graphics Forum* (2020), vol. 39, Wiley Online Library, pp. 623–642. 5
- [WGSJ20] WILES O., GKIOXARI G., SZELISKI R., JOHNSON J.: Synsin: End-to-end view synthesis from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 7467–7477. 3
- [WSMG*16] WHELAN T., SALAS-MORENO R. F., GLOCKER B., DAVISON A. J., LEUTENEGGER S.: Elasticfusion: Real-time dense slam and light source estimation. *The International Journal of Robotics Research* 35, 14 (2016), 1697–1716. 3
- [XXP*22] XU Q., XU Z., PHILIP J., BI S., SHU Z., SUNKAVALLI K., NEUMANN U.: Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), pp. 5438–5448. 3
- [YCA*20] YANG Z., CHAI Y., ANGUELOV D., ZHOU Y., SUN P., ERHAN D., RAFFERTY S., KRETZSCHMAR H.: Surfelgan: Synthesizing realistic sensor data for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2020). 3
- [YLT*21] YU A., LI R., TANCİK M., LI H., NG R., KANAZAWA A.: Plenotrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 5752–5761. 3
- [YLY*19] YU J., LIN Z., YANG J., SHEN X., LU X., HUANG T. S.: Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (2019), pp. 4471–4480. 4
- [YSW*19] YIFAN W., SERENA F., WU S., ÖZTIRELI C., SORKINE-HORNUNG O.: Differentiable surface splatting for point-based geometry processing. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–14. 3
- [YYTK21] YU A., YE V., TANCİK M., KANAZAWA A.: pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 4578–4587. 2
- [ZBRH22] ZHANG Q., BAEK S.-H., RUSINKIEWICZ S., HEIDE F.: Differentiable point-based radiance fields for efficient view synthesis. *arXiv preprint arXiv:2205.14330* (2022). 3
- [ZIE*18] ZHANG R., ISOLA P., EFROS A. A., SHECHTMAN E., WANG O.: The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR* (2018). 6
- [ZPVBG01] ZWICKER M., PFISTER H., VAN BAAR J., GROSS M.: Surface splatting. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), pp. 371–378. 3
- [ZTF*18] ZHOU T., TUCKER R., FLYNN J., FYFFE G., SNAVELY N.: Stereo magnification: Learning view synthesis using multiplane images. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–12. 2
- [ZTS*16] ZHOU T., TULSIANI S., SUN W., MALIK J., EFROS A. A.: View synthesis by appearance flow. In *European conference on computer vision* (2016), Springer, pp. 286–301. 2