

# PARAMETRIC SDF FOR DYNAMIC SURFACE RECONSTRUCTION

Anonymous authors

Paper under double-blind review

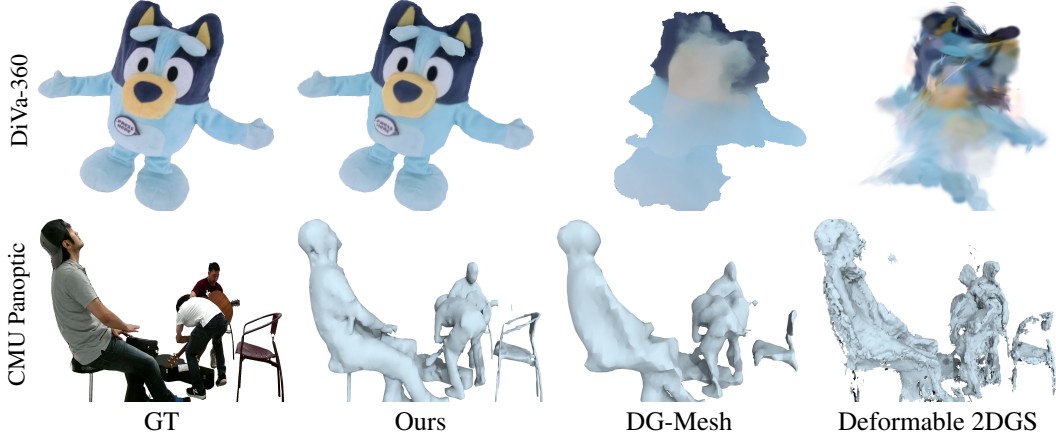


Figure 1: p-SDF produces high-fidelity dynamic 3D surface reconstructions. Compared with prior arts (Liu et al., 2024; Zhang et al., 2024), our method preserves more faithful appearance and geometry across challenging real-world dynamic datasets (Joo et al., 2015; Lu et al., 2024).

## ABSTRACT

Reconstructing high-fidelity, temporally coherent surfaces of dynamic scenes remains a critical challenge in computer vision. While recent methods excel at novel view synthesis, they often fail to recover accurate geometry, yielding noisy or temporally inconsistent meshes that are suboptimal for downstream applications such as simulation or editing. In this work, we introduce a new paradigm for dynamic surface reconstruction based on a parametric Signed Distance Function (p-SDF). Our key insight is to generalize static SDF fields—where each spatial point stores a constant value—into time-dependent parametric curves with each curve modeling a temporally evolving SDF trajectory. Such a parametric SDF modeling provides a principled way to capture complex temporal variations, naturally enforcing smoothness and continuity in shape dynamics. At each timestamp, a static SDF field can be queried from p-SDF and converted into an explicit surface mesh via differentiable iso-surfacing. By rendering these meshes with a physically based differentiable renderer, we optimize the underlying parametric curves end-to-end against 2D image observations. Our framework produces high-fidelity, temporally coherent surfaces and inherently disentangles geometry, material, and lighting from multi-view videos. It robustly reconstructs geometry under large-scale motions and resolves appearance ambiguities caused by challenging lighting and occlusions. Experiments on both synthetic and real-world scenes demonstrate that our method achieves state-of-the-art geometric accuracy and temporal consistency, delivering delicate meshes that surpass prior work, benefiting from our parametric SDF representation.

## 1 INTRODUCTION

Reconstructing high-fidelity surfaces from multi-view video is a long-standing pursuit in computer vision, powering applications in simulation, virtual and augmented reality (VR/AR), autonomous robotics, and film production. A central requirement is *temporal coherence*: the reconstructed geometry must evolve smoothly and consistently over time. Without it, surfaces may flicker, distort, or break apart—undermining downstream tasks such as relighting, physical simulation, or editing.

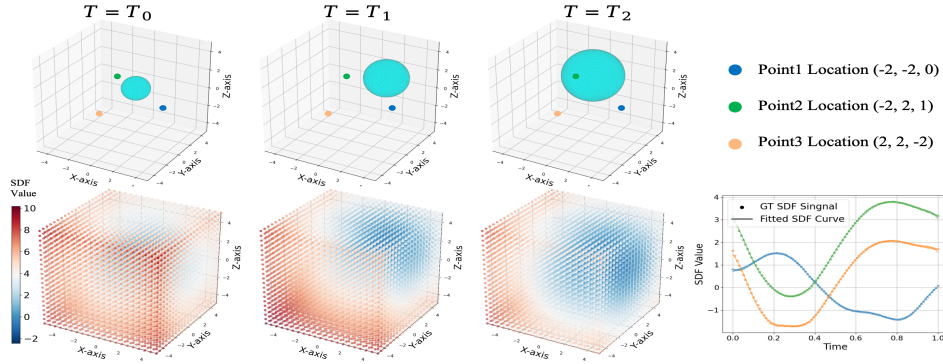


Figure 2: **Continuous SDF Trajectory.** The SDF field of a dynamic scene evolves over time, e.g., a sphere dilating and moving (left top) cause the variations of the SDF field (left bottom). Each spatial point (e.g., green, orange, and blue markers) traces a continuous SDF trajectory. We optimize our p-SDF with a compact set of basis functions (Sec. 3.3) to represent these continuous curves. As shown in the right figure, p-SDF closely follows the ground truth SDF trajectories.

Achieving this coherence, however, remains highly challenging, as continuous deformations of dynamic scenes often involve non-rigid motion, large-scale dynamics, and even topological changes.

Recent advances in scene representations, such as Neural Radiance Fields (NeRFs) (Mildenhall et al., 2021) and 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023), have dramatically advanced novel view synthesis (NVS). Their dynamic extensions (Pumarola et al., 2021; Fridovich-Keil et al., 2023; Cao & Johnson, 2023; Yang et al., 2024; Duan et al., 2024) further enable photorealistic NVS across time. Nevertheless, these methods are designed primarily for appearance modeling and thus struggle with geometry, often producing noisy or temporally unstable surfaces. Two lines of work have explored improving quality for dynamic surface reconstruction. The first strategy learns a deformation field that maps a canonical template to the dynamic scene over time (Liu et al., 2024; Zhang et al., 2024). While this strategy enforces temporal consistency, it is inherently limited by the fixed template, which fails to represent topological changes. The second strategy directly models the 4D spatio-temporal volume through factorized representations such as Tri-planes or Hash grids (Wang et al., 2023; Chen et al., 2025; Zheng et al., 2025). Although geometry can be extracted by querying the 4D volume, these methods fail to maintain temporal coherence, often producing noisy or flickering reconstructions under complex dynamics or large motions.

In this work, we propose parametric Signed Distance Function (p-SDF), a novel dynamic scene representation that reconstructs high-fidelity, temporally coherent surfaces across time. Our key intuition is that most shape deformations are continuous along time, and the evolution of SDF values at a particular 3D location follows a continuous trajectory (as shown in a toy example in Fig. 2, more detailed analysis in Appendix A). This observation motivates us to generalize static SDF fields, where each spatial point stores a constant scalar value, into time-dependent parametric curves, where temporally evolving SDF trajectory is parametrized with a curve defines. By using a compact set of basis functions to represent these parametric curves, our p-SDF naturally enforces the temporal coherence of the geometry while being flexible in capturing complex dynamics, including large deformations and topological changes.

Specifically, we represent the dynamic geometry with a grid where each vertex stores the parameters of the basis functions. At any timestamp, we can directly query the curve to obtain a static SDF field, from which an explicit mesh can be extracted by differentiable iso-surfacing (Shen et al., 2023). The scene’s appearance is modeled separately with an efficient 4D hash grid that predicts physically based (PBR) material properties. The resulting mesh and materials are then passed to a differentiable renderer, allowing the entire framework to be optimized end-to-end against 2D image observations. Our method offers two key advantages: (i) p-SDF directly models the variations of SDF fields, enforcing temporal coherence for geometry while allowing topological changes and large-scale motions; (ii) by disentangling geometry from a physically based appearance modeling, our framework robustly reconstructs surfaces under challenging lighting and occlusion conditions.

Our p-SDF presents a principled and efficient solution for dynamic 3D reconstruction. We evaluate our method on synthetic and real-world datasets, demonstrating significant performance improvement over prior approaches in geometric accuracy and temporal consistency. Moreover, the high-quality material decomposition from our framework readily supports relighting applications, enabling physically accurate relighting effects.

## 2 RELATED WORK

### 2.1 NEURAL REPRESENTATIONS FOR DYNAMIC SCENES

The advent of Neural Radiance Fields (NeRF) (Mildenhall et al., 2021) revolutionized static novel view synthesis by representing a scene as a continuous volumetric function learned from a collection of 2D images. Extending this paradigm to dynamic scenes involved modeling a 4D space-time function, often by incorporating time-dependent deformations or a time-variant latent code to map observations to a canonical space (Du et al., 2021; Park et al., 2021a,b; Pumarola et al., 2021; Fang et al., 2022; Guo et al., 2023; Yan et al., 2024; Herau et al., 2024; Kumar et al., 2025).

A more recent paradigm shift in neural rendering came with the introduction of 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023), which replaced the implicit volumetric function with a collection of explicit 3D Gaussians. This approach enables state-of-the-art visual quality at real-time rendering speeds and has quickly become the new frontier for dynamic scene modeling. Typically, these methods learn a deformation field that maps Gaussians from a canonical space to the observed space (Huang et al., 2023; Wu et al., 2024; Xu et al., 2024; Ren et al., 2024; Yang et al., 2024; Wang et al., 2025), or introduce elaborate data structures to deform the dense set of Gaussians (Duan et al., 2024; Zhu et al., 2025; Wu et al., 2025). However, Gaussian primitives combined with deformation fields lack strong geometric priors: while effective for view synthesis, they do not enforce temporal coherence and often yield noisy, incomplete, or temporally inconsistent meshes (Cai et al., 2024; Chen et al., 2025; Zhang et al., 2024).

### 2.2 DYNAMIC SURFACE RECONSTRUCTION

A distinct branch of work explicitly targets surface reconstruction quality, typically relying on Signed Distance Function (SDF) representations and their variants (Wang et al., 2021; Shen et al., 2021, 2023; Munkberg et al., 2022). Their extensions to dynamic reconstruction often adopt deformation field modeling, where a single high-quality canonical template surface is learned and deformed over time (Cai et al., 2022; Johnson et al., 2023; Wang et al., 2024; Yao et al., 2024). While this formulation yields temporally stable and smooth surfaces, it inherently assumes fixed topology, making it incapable of handling topological changes in many dynamic scenarios.

Alternatively, a body of work directly models the 4D spatio-temporal volume, thereby avoiding topological constraints (Niemeyer et al., 2019; Choe et al., 2023; Mao et al., 2024; Liu et al., 2024; Sang et al., 2025). These methods typically define a unique implicit surface at each timestamp and recover explicit meshes using iso-surfacing techniques (Ju et al., 2002; Chen et al., 2022; Shen et al., 2021, 2023). However, they lack a strong temporal prior: without canonical guidance, each surface is reconstructed with excessive degrees of freedom, often leading to flickering, jittering, or other temporal artifacts that require heavy regularization.

Similar to our method, SDFFlow (Mao et al., 2024) also models the deformation of SDF fields. However, its integral formulation leads to inefficient training and accumulated error, making it impractical for longer sequences and fail to achieve high-quality reconstruction. In contrast, our method directly parameterizes the SDF trajectories with basis functions, inherently enforcing continuity regularization. The static SDF at a given timestamp can be obtained efficiently by querying the function, without explicit integration that is extremely time-consuming and error-accumulating.

## 3 METHODOLOGY

We now describe our method. The key idea is to generalize static Signed Distance Function (SDF) fields into parametric SDF trajectories. Below, we first review background concepts on SDF and temporal signal modeling in Sec. 3.1. We then provide an overview of the full framework in Sec. 3.2 followed by detailed descriptions of our parametric SDF formulation (Sec. 3.3) and appearance modeling (Sec. 3.4). Lastly, we talk about the implementation details in Sec. 3.5.

### 3.1 BACKGROUND

Our approach integrates principles from explicit surface reconstruction and temporal signal modeling with basis functions. We first provide preliminary knowledges of these two concepts.

**Surface Reconstruction with Signed Distance Functions.** A 3D surface  $\partial\Omega$  can be represented as the zero-level set of a Signed Distance Function (SDF). For any spatial point  $\mathbf{x} \in \mathbb{R}^3$ , the SDF value

$s(\mathbf{x})$  is the signed distance to  $\partial\Omega$ , negative inside the enclosed volume  $\Omega$  and positive outside:

$$s(\mathbf{x}) = \begin{cases} -\inf_{\mathbf{p} \in \partial\Omega} \|\mathbf{x} - \mathbf{p}\|_2 & \text{if } \mathbf{x} \in \Omega, \\ \inf_{\mathbf{p} \in \partial\Omega} \|\mathbf{x} - \mathbf{p}\|_2 & \text{otherwise.} \end{cases} \quad (1)$$

Explicit surface reconstruction methods (Munkberg et al., 2022) define SDF values on a 3D grid and use differentiable iso-surfacing (Shen et al., 2021; 2023) to extract a triangle mesh. The resulting mesh can then be rendered via a differentiable rasterizer (Laine et al., 2020). While highly effective for static scenes, naively applying such pipelines into dynamics, *i.e.*, applying NVDiffRec via per-frame reconstruction strategy, is computationally expensive and does not guarantee temporal consistency, often resulting in flickering geometry and unstable appearance. Instead, our method develops a unified 4D surface representation that preserves the advantages of explicit meshes while enforcing smooth surface evolution over time.

**Temporally Varying Signal Modeling with Basis Functions.** A standard strategy for representing a temporal signal  $S(t)$  is to express it as a linear combination of basis functions,  $S(t) = \sum_{k=1}^K w_k B_k(t)$ , where  $\{B_k\}_{k=1}^K$  are predefined temporal basis and  $\{w_k\}_{k=1}^K$  are learnable coefficients. With appropriate choices for basis, a small set of coefficients compactly encodes complex temporal behavior while naturally encouraging smoothness.

The common options of basis include Polynomial functions or Fourier basis functions. Specifically, Polynomial basis functions uses a combination of polynomial functions:

$$B_k^p(t) = t^k \quad (2)$$

which captures smooth and low-frequency trends. Alternatively, Fourier basis functions leverages a combinations of periodic functions based on  $\sin, \cos$ :

$$B_k^f(t) = a_k \cos(\omega_k t) + b_k \sin(\omega_k t), \quad (3)$$

where the  $\omega_k$  is the predefined frequency for  $k$ -th periodic function, and  $a_k$  and  $b_k$  is the coefficients. The Fourier basis foundations are well-suited for periodic or high-frequency variations as shown in literature (Shi et al., 2014; Tancik et al., 2020; Wang et al., 2022; Rabich et al., 2024; Li et al., 2024).

### 3.2 FRAMEWORK OVERVIEW

**Motivation.** Representing dynamic scenes requires modeling how surfaces evolve over time. As illustrated in Fig. 2 our key observation is that surface motions induces smooth temporal evolution of the SDF values at any fixed spatial location, *i.e.*,  $s(\mathbf{x}, t)$  traces a continuous 1D signal. These continuous 1D signals can be efficiently parameterized via a collection of basis functions, which not only offers interpretability of the SDF trajectories, but also naturally enforces temporal coherence in the SDF evolution while retaining flexibility for large deformations and topological changes.

**Pipeline.** Fig. 3 overviews the full system. Given a timestamp  $t$ , we first obtain per-vertex SDF values via evaluating basis functions and extracts an explicit mesh using differentiable iso-surfacing (Shen et al., 2023). Concurrently, to model the appearance, we apply a 4D Hash grid and query the PBR attributes for each vertex in the mesh. A physically based differentiable renderer then synthesizes the target views; gradients backpropagate through geometry, appearance, and rendering, enabling end-to-end optimization (Sec. 3.5). This design yields temporally coherent, high-quality meshes and clean disentanglement of geometry, materials, and lighting.

### 3.3 PARAMETRIC SIGNED DISTANCE FUNCTION

**Formulation.** We now introduce our dynamic geometry representation, p-SDF, which extends static SDF fields into the temporal domain. Specifically, the scene geometry is defined within a bounded 3D grid  $G$  with  $N \times N \times N$  vertices  $\{\mathbf{v}_i\}_{i=1}^{N^3}$ . Unlike static methods that assign a single SDF value to each vertex, we model the SDF value at vertex  $\mathbf{v}_i$  as a continuous function of time:  $s(\mathbf{v}_i, t) : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}$ . Following prior work on temporal signal modeling, we adopt a hybrid basis combining polynomials and Fourier terms to represent  $s(\mathbf{v}_i, t)$ :

$$s(\mathbf{v}_i, t) = \sum_{n=1}^{N_p} \mu_n t^n + a_0 + \sum_{n=1}^{N_f} (a_n \cos(\omega_n t) + b_n \sin(\omega_n t)), \quad (4)$$



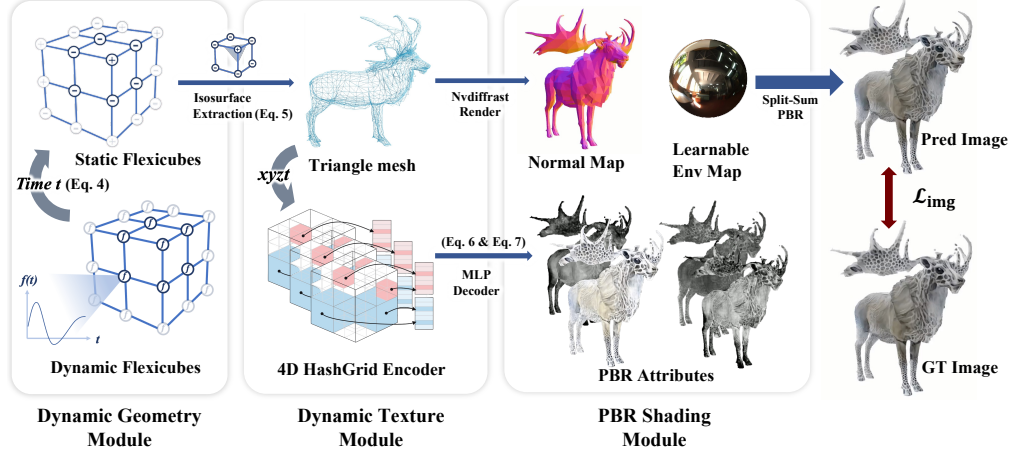


Figure 3: **Pipeline.** For a timestep  $t$ , p-SDF query the parametric function at each grid vertex to get the SDF value (Eq. 4). We use FlexiCubes (Shen et al., 2023) to extract the mesh  $M_t = (V_t, F_t)$  (Eq. 5), and query the Dynamic Tri-plane Appearance Field (Sec. 3.4) to obtain  $(c, m, r, o)$  (Eq. 7) at point  $(x, t)$ . A differentiable PBR shader renders images from multi-view cameras, and gradients propagate through all modules for end-to-end training (Sec. 3.5).

where  $N_p$  is the number of polynomials, and  $N_f$  is the highest frequency in Fourier function. Collectively, the entire spatio-temporal SDF field is compactly parameterized by  $W \in \mathbb{R}^{N^3 \times (N_p + 2N_f + 1)}$ , the matrix of all vertex coefficients.

**Mesh Extraction.** With our time-varying SDF grids, given a timestamp  $t$ , we can evaluate the basis function to get per-vertex SDF values. A high-quality, watertight surface mesh is then extracted using differentiable Dual Marching Cubes (Shen et al., 2023):

$$M_t = \text{DualMarchingCubes}(\{s(\mathbf{v}_i, t)\}_i^{N^3}), \quad M_t := (V_t, F_t), \quad (5)$$

where  $V_t$  and  $F_t$  denote the mesh vertices and faces at time  $t$ .

**Discussion.** This parametric design for surface mesh provides several important benefits; (i) the basis formulation inherently enforces temporal coherence, while the combination of polynomial and Fourier bases allows capturing both large-scale deformations and fine-grained high-frequency motions, offering robustness in complex dynamic scenes; (ii) we model the geometry by storing per-vertex curve parameters on a grid, which enables efficient and scalable mesh extraction by directly querying the stored function, thus bypassing the computational overhead of per-point MLPs in prior works. (iii) by using the explicit mesh for rendering within the differentiable rendering-optimization loop, we apply direct supervision to the geometry. This ensures a consistent and detailed surface, avoiding the noisy artifacts common in purely implicit methods like NeuS (Wang et al., 2023).

### 3.4 APPEARANCE REPRESENTATION

**Formulation.** Having established a continuous representation for dynamic geometry, we now model the time-varying appearances and materials. Unlike geometry, appearance signals such as RGB textures often undergo abrupt changes at fixed spatial locations and therefore cannot be compactly represented by smooth temporal trajectories. To address this, we separately model the appearance with a temporal tri-plane representation, parameterizing the appearance field with three multi-resolution 3D hash grids,  $G_{xyt}$ ,  $G_{xzt}$ , and  $G_{yzt}$ , which cover the spatio-temporal domains  $(x, y, t)$ ,  $(x, z, t)$ , and  $(y, z, t)$ . For a surface point  $(x, y, z)$  at time  $t$ , we query these grids in parallel via trilinear interpolation:

$$\mathbf{f}_{xyt} = G_{xyt}(x, y, t), \quad \mathbf{f}_{xzt} = G_{xzt}(x, z, t), \quad \mathbf{f}_{yzt} = G_{yzt}(y, z, t). \quad (6)$$

The resulting feature vectors are concatenated and decoded by a lightweight MLP  $\Phi_{\text{app}}$  into physically based material properties:

$$(c, m, r, o) = \Phi_{\text{app}}(\text{concat}(\mathbf{f}_{xyt}, \mathbf{f}_{xzt}, \mathbf{f}_{yzt})), \quad (7)$$

where  $c$  denotes the diffuse albedo,  $m$  the metallic value,  $r$  the surface roughness, and  $o$  the self-occlusion attribute. At any timestamp  $t$ , this formulation defines a complete, fine-grained PBR texture field aligned with the mesh  $M_t$  extracted from the p-SDF.

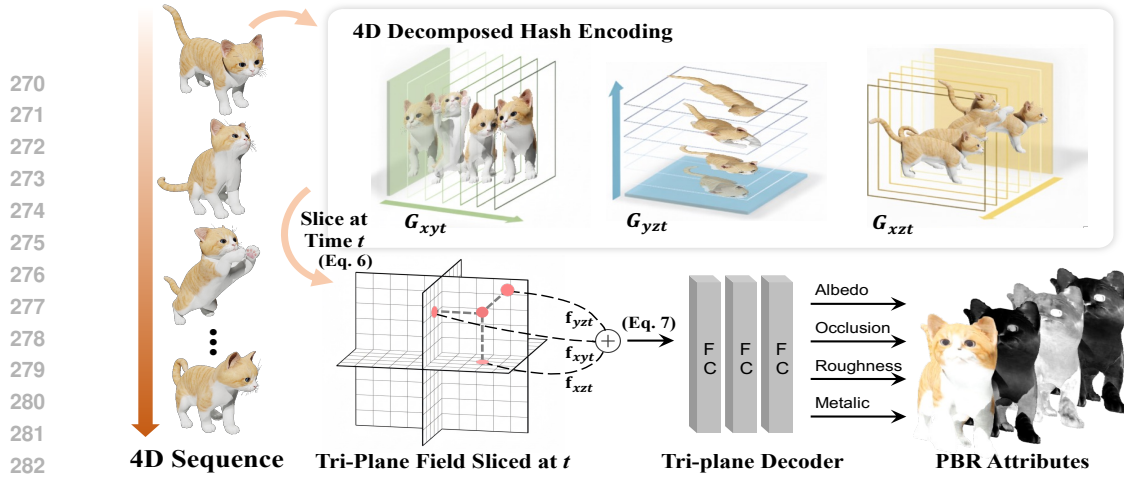


Figure 4: **Dynamic Tri-plane Appearance Field.** Left: an input multi-view video (4D sequence). Middle: three multi-resolution spatio-temporal hash grids  $G_{xyt}, G_{xzt}, G_{yzt}$  that slice at time  $t$  to produce an instantaneous tri-plane. Right: features sampled from the three planes are concatenated and fed into a small MLP decoder that predicts PBR attributes (albedo  $c$ , occlusion  $o$ , roughness  $r$ , and metallic  $m$ ) for shading and differentiable rendering.

**Rendering.** At time  $t$ , with the extracted mesh  $M_t = (V_t, F_t)$  from the geometry module (Sec. 3.3), we first rasterize it and generate a G-buffer of world-space positions  $\mathbf{x}$  and surface normals  $\mathbf{n}$ . For each visible surface point, the Dynamic Tri-plane Appearance Field is queried at  $(\mathbf{x}, t)$  to produce material properties  $(c, m, r, o)$  (Eq. 7). These attributes, combined with the mesh geometry and a shared learnable environment map, are passed to a differentiable PBR shader, following Nvdiffrec (Munkberg et al., 2022), to compute the final rendered RGB image. This design enables end-to-end optimization of both geometry and appearance against multi-view video observations.

**Discussion.** Our appearance representation offers two advantages: (i) by decomposing the 4D spatio-temporal domain into three 3D hash grids, it reduces computational complexity to directly model 4D appearance; (ii) Benefiting from the precise geometry, the simple approach suffices to achieve a clean disentanglement of shape and appearance and produce high-quality material and lighting estimation.

### 3.5 IMPLEMENTATION DETAILS

**Loss Function.** We employ a compact objective that balances photometric reconstruction with geometric and appearance priors. The overall loss is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{photo}} + \mathcal{L}_{\text{curve}} + \mathcal{L}_{\text{app}}, \quad (8)$$

where  $\mathcal{L}_{\text{photo}}$  enforces consistency between rendered and ground-truth images,  $\mathcal{L}_{\text{curve}}$  regularizes the spatial smoothness of SDF trajectories to ensure coherent motion for neighboring vertices, and  $\mathcal{L}_{\text{app}}$  regularizes material and lighting. Detailed formulations of these losses are provided in Appendix B.

**Model Details.** We use a grid of resolution  $96^3$  to represent our p-SDF, and use a combination of 6 polynomial basis functions and a set of Fourier components, the number of which is chosen per-scene from the range  $[18, 100]$  based on motion complexity. For the Dynamic Tri-plane Appearance Field, we adapt the multi-resolution hash grid following Grid4D (Xu et al., 2024), with spatial resolution spanning from 16 to 2048 and temporal resolution from 8 to the full sequence length.

**Training and Optimization.** We build our framework upon the open-source RadianceFieldStudio codebase (Ye et al., 2024) and conduct all experiments on a single NVIDIA RTX 4090 GPU. We train each scene for 5,000 iterations, which takes approximately 1 to 3 hours. We optimize the model using an Adam optimizer (Kingma, 2014) with exponential decay.

## 4 EXPERIMENTS

### 4.1 EXPERIMENTAL SETUP

**Datasets.** To rigorously validate our method, we evaluate on both synthetic and real-world datasets that feature challenging dynamic scenes. Specifically, we select three datasets:

Table 1: Quantitative evaluation of our method compared to previous work on *SynthoMotion-360* and *DiVa-360* datasets. Our method outperforms existing approaches in terms of both visual fidelity and geometric quality. For clarity, all Chamfer distances are scaled by  $10^{-3}$ .

Method	SynthoMotion-360					DiVa-360		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	MAE $\downarrow$	Chamfer $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Deformable-3DGS	28.897	0.956	0.077	2.842	4.535	27.280	0.956	0.068
SC-GS	29.881	0.962	0.058	2.372	3.051	25.723	0.951	0.075
Dynamic-2DGS	29.489	0.959	0.062	2.461	2.850	17.513	0.910	0.187
Grid4D	30.048	0.962	0.061	2.639	4.637	27.408	0.954	0.065
DG-Mesh	25.719	0.936	0.081	2.794	3.760	17.734	0.919	0.183
Ours	<b>31.653</b>	<b>0.974</b>	<b>0.027</b>	<b>1.332</b>	<b>2.220</b>	<b>28.076</b>	<b>0.957</b>	<b>0.049</b>

*SynthoMotion-360*. We first curate a *new synthetic benchmark* designed to assess reconstruction under large, non-rigid deformations that existing datasets rarely capture. While existing datasets primarily feature constrained human or object motions, ours introduces a diverse range of subjects, including animals and animated characters, undergoing extreme non-rigid deformations. These challenging scenarios are absent in current benchmarks. In particular, we carefully select 7 animated assets from Objaverse (Deitke et al., 2022; Liang et al., 2024), each with 70–300 frames rendered in Blender from 38 cameras. We use 32 views for training and 6 for testing. Accurate ground-truth meshes are provided for every frame, enabling quantitative evaluation of both novel view synthesis quality and geometry accuracy. We provide overview of our dataset in Appendix E.

*DiVa-360* is a real-world multi-view dataset featuring tabletop dynamic scenes captured under synchronized RGB cameras (Lu et al., 2024). It provides synchronized videos and foreground masks but no ground-truth meshes, so we only evaluate photometric fidelity in this dataset. We use 4 representative sequences, each using 31 cameras for training and 6 held-out views for evaluation.

*CMU Panoptic Studio* is a representative real-world dynamic dataset featuring complex motions (Joo et al., 2015). Following the protocol of SDFFlow (Mao et al., 2024), we focus on challenging multi-person interaction sequences, each containing 24 frames recorded from 10 calibrated cameras. As this dataset lacks ground-truth meshes, we use it for qualitative comparison to demonstrate the quality of geometry. We use all 10 views for training.

**Metrics.** We evaluate the performance from two aspects: *novel-view synthesis quality* and *geometric reconstruction accuracy*. For novel view synthesis quality, we report PSNR, SSIM (Wang et al., 2004), and LPIPS scores (Zhang et al., 2018). For geometric reconstruction quality, we report the Chamfer-L1 distance and the Mean Angular Error (MAE) of surface normals, following the protocol of Verbin et al. (2022).

**Baselines.** We benchmark our method against state-of-the-art baselines from two representative research directions: *dynamic surface reconstruction* and *dynamic novel view synthesis*. For the first one, we compare with two state-of-the-art works: Dynamic-2DGS (Zhang et al., 2024) and DG-Mesh (Liu et al., 2024), which extracts consistent meshes from dynamic 3D Gaussians. We note that while SDFFlow (Mao et al., 2024) is a conceptually related SDF-based approach, it is excluded due to its prohibitive optimization time, requiring several weeks per scene. For the second one, we compare with three state-of-the-art works: Deformable-3DGS (Yang et al., 2024), SC-GS (Huang et al., 2023) and Grid4D (Xu et al., 2024). This selection allows a comprehensive evaluation of our model’s superiority in both geometric accuracy and novel view synthesis quality.

## 4.2 EXPERIMENTAL RESULTS

**Quantitative Results.** We provide quantitative comparisons with all the baselines in Tbl. I with detailed results of each scene in Appendix Tbl A1 and Tbl A2. Our approach consistently outperforms all the competing methods across all the metrics on both the synthetic *SynthoMotion-360* and the real-world *DiVa-360* datasets, demonstrating the superiority in reconstructing both geometry and visual appearance. In particular, on *SynthoMotion-360* benchmark, we reduce the MAE by over 40% and Chamfer Distance by over 20% compared to the best-performing baseline. This improvement highlights the core advantage of our p-SDF representation, which inherently enforces temporal consistency and is well suited for capturing complex non-rigid dynamics. With p-SDF providing precise and consistent geometry, the dynamic tri-plane appearance field further produces faithful material decomposition, rendering images that have much better visual quality than rendering-focused state-of-the-art methods without requiring specialized optimizations for appearance modeling.

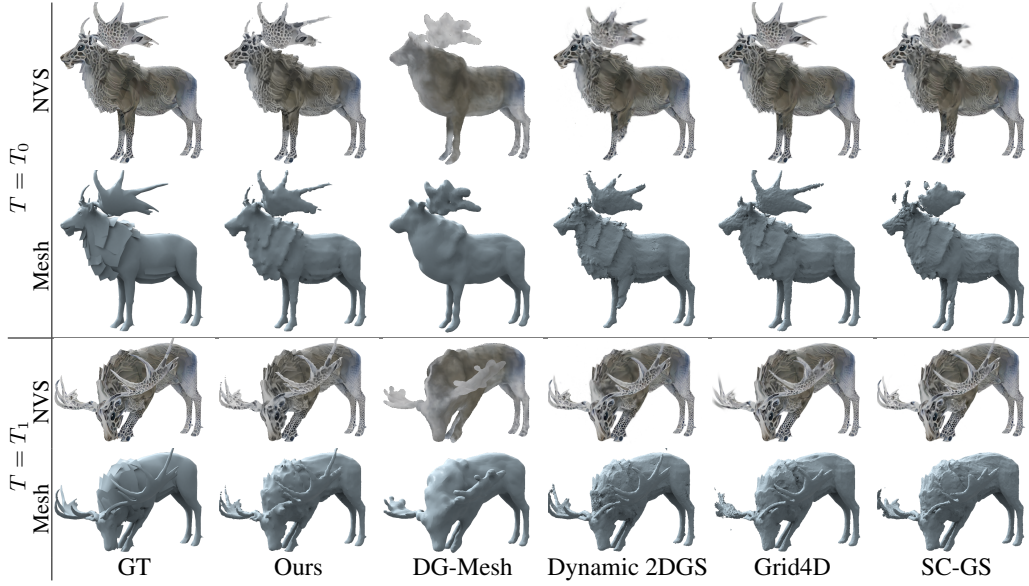


Figure 5: **Qualitative comparison on SynthoMotion360 dataset.** We compare against baselines on both visual (visualized as shaded meshes) and geometry quality. Our method reconstructs realistic appearance and accurate geometry with fewer artifacts for large motions.

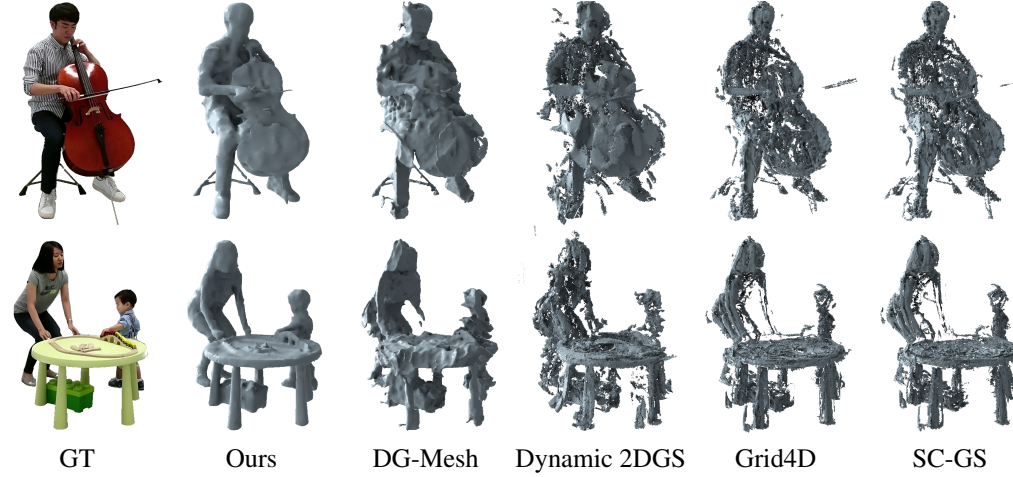


Figure 6: **Qualitative comparison on CMU Panoptic Studio dataset.** We compare the geometry quality against all baselines. Our method reconstructs accurate geometry, while baselines for surface reconstruction produce noticeable artifacts and broken geometry.

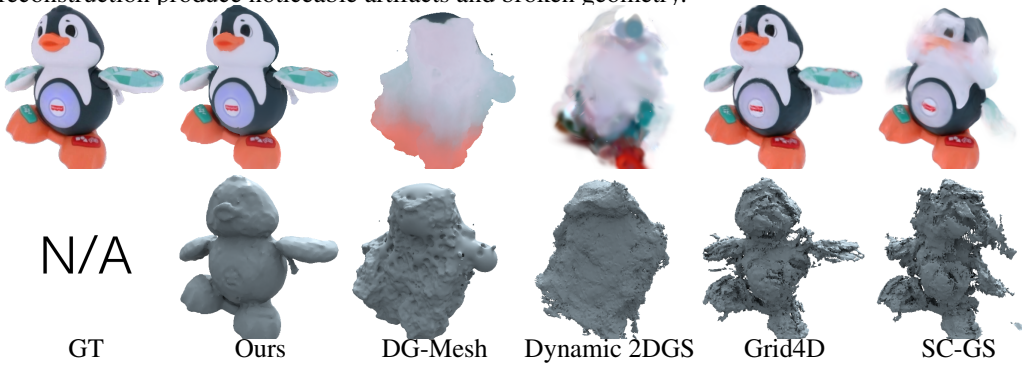


Figure 7: **Qualitative comparison on DiVa-360 dataset.** Similarly, our method reconstructs accurate geometry, while baselines for surface reconstruction and dynamic NVS all produce noticeable artifacts and broken geometry.



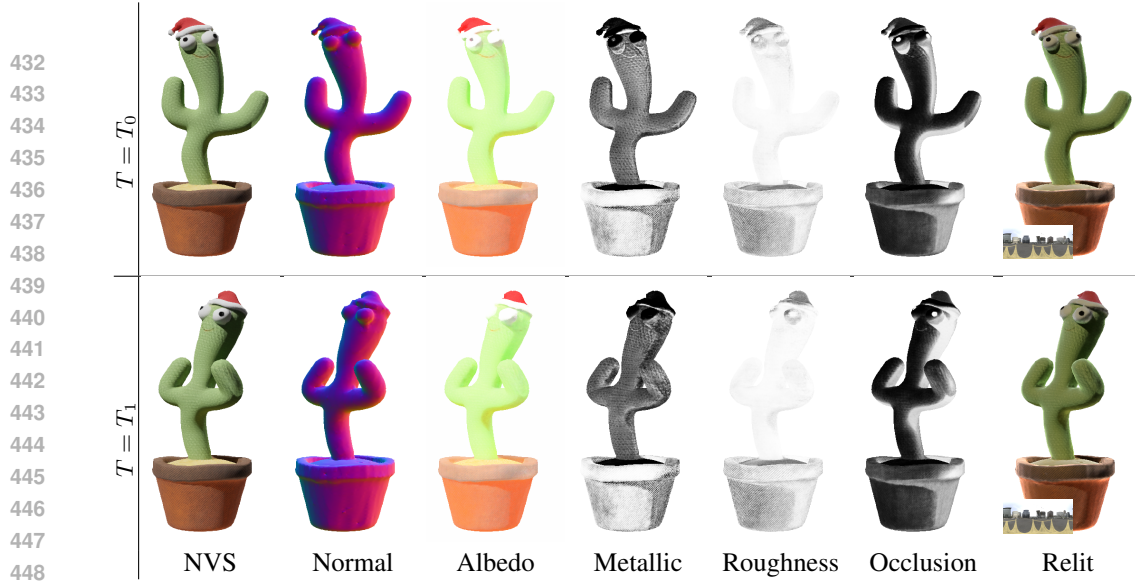


Figure 8: **PBR material reconstruction and relighting results.** Our method reconstructs plausible PBR materials (albedo, metallic and roughness) from multi-view videos, supporting re-lighting applications with new environment maps.

**Qualitative Results.** We present qualitative comparisons in Fig. 1, 5, 6 and 7, showing results from *SynthaMotion360*, *CMU Panoptic Studio* and *DiVa-360*, respectively. Additional results are provided in Appendix Fig. A3, A4, A5 and A6. In geometric reconstruction, the surfaces recovered by baselines are severely flawed, suffering from fragmented geometry, incompleteness and noise. For example, DG-Mesh, despite being optimized for surface quality, produces results with noticeable artifacts and distortions. In contrast, our method reconstructs high-fidelity surfaces that are complete, smooth, and topologically sound, accurately capturing fine details such as the structure of the cello and the distinct figures of the people in Fig. 6. In appearance, the geometric deficiencies of baselines directly impact rendering quality of baselines. They produce temporally inconsistent results (the antlers in Fig. 5) or as overly smoothed, blurry renderings (the penguin in Fig. 7). In contrast, benefiting from the precise and stable geometry learned by our p-SDF, our method produces sharp, detailed, and photorealistic images, faithfully synthesizing fine textures and maintaining temporal coherence even during complex dynamic sequences.

#### 4.3 RELIGHTING

To further validate the accuracy and reliability of our dynamic reconstruction, we conduct relighting experiments using our reconstructed geometry and PBR material. The results are shown in Fig. 8. The reconstructed results enable high-quality relighting for dynamic scenes. To the best of our knowledge, our method is the first to achieve dynamic reconstructions with precise PBR materials to support realistic dynamic relighting.

## 5 CONCLUSIONS AND LIMITATIONS

**Conclusions.** In this paper, we introduced parametric Signed Distance Function (p-SDF), a novel dynamic scene representation that leverages parametric basis functions to represent SDF trajectories. p-SDF naturally enforces temporally coherent geometry while remaining flexible to large deformations and topological changes. Combined with a physically based appearance modeling, our framework yields accurate geometry, robust material decomposition in highly dynamic conditions, and effectively supports downstream application such as relighting. Experiments on both synthetic and real-world datasets demonstrate substantial improvements over prior methods in geometric fidelity and visual quality. We believe p-SDF offers a promising direction for dynamic surface reconstruction and its applications.

**Limitations.** Our approach relies on multi-view video as input, which may limit its applicability in scenarios where only monocular data is available. In addition, the parametric SDF we used leverages the basis functions at each fixed spatial location, and thus, it does not explicitly represent the correspondence information along the time.

## REFERENCES

- Hongrui Cai, Wanquan Feng, Xuetao Feng, Yan Wang, and Juyong Zhang. Neural surface reconstruction of dynamic scenes with monocular rgb-d camera. *Advances in Neural Information Processing Systems*, 35:967–981, 2022.
- Weiwei Cai, Weicai Ye, Peng Ye, Tong He, and Tao Chen. Dynasurfgs: Dynamic surface reconstruction with planar-based gaussian splatting. *arXiv preprint arXiv:2408.13972*, 2024.
- Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 130–141, 2023.
- Decai Chen, Brianne Oberson, Ingo Feldmann, Oliver Schreer, Anna Hilsmann, and Peter Eisert. Adaptive and temporally consistent gaussian surfels for multi-view dynamic reconstruction. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 742–752. IEEE, 2025.
- Zhiqin Chen, Andrea Tagliasacchi, Thomas Funkhouser, and Hao Zhang. Neural dual contouring. *ACM Transactions on Graphics (TOG)*, 41(4):1–13, 2022.
- Jaesung Choe, Christopher Choy, Jaesik Park, In So Kweon, and Anima Anandkumar. Spacetime surface regularization for neural dynamic scene reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17871–17881, 2023.
- Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects, 2022. URL <https://arxiv.org/abs/2212.08051>
- Yilun Du, Yinan Zhang, Hong-Xing Yu, Joshua B Tenenbaum, and Jiajun Wu. Neural radiance flow for 4d view synthesis and video processing. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14304–14314. IEEE Computer Society, 2021.
- Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d-rotor gaussian splatting: Towards efficient novel-view synthesis for dynamic scenes. In *Proc. SIGGRAPH*, July 2024.
- Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–9, 2022.
- Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12479–12488, 2023.
- Xiang Guo, Jiadai Sun, Yuchao Dai, Guanying Chen, Xiaoqing Ye, Xiao Tan, Errui Ding, Yumeng Zhang, and Jingdong Wang. Forward flow for novel view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16022–16033, 2023.
- Quentin Herau, Nathan Piasco, Moussab Bennehar, Luis Roldao, Dzmitry Tsishkou, Cyrille Migniot, Pascal Vasseur, and Cédric Demonceaux. Soac: Spatio-temporal overlap-aware multi-sensor calibration using neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15131–15140, 2024.
- Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. *arXiv preprint arXiv:2312.14937*, 2023.
- Erik Johnson, Marc Habermann, Soshi Shimada, Vladislav Golyanik, and Christian Theobalt. Unbiased 4d: Monocular 4d reconstruction with a neural deformation model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6598–6607, 2023.

- Hanbyul Joo, Hao Liu, Lei Tan, Lin Gui, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for social motion capture. In *Proceedings of the IEEE international conference on computer vision*, pp. 3334–3342, 2015.
- Tao Ju, Frank Losasso, Scott Schaefer, and Joe Warren. Dual contouring of hermite data. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pp. 339–346, 2002.
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.
- Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Ashish Kumar et al. Dynamode-nerf: Motion-aware deblurring neural radiance field for dynamic scenes. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 21728–21738, 2025.
- Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. Modular primitives for high-performance differentiable rendering. *ACM Transactions on Graphics*, 39(6), 2020.
- Zhengqi Li, Richard Tucker, Noah Snavely, and Aleksander Holynski. Generative image dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.
- Hanwen Liang, Yuyang Yin, Dejia Xu, Hanxue Liang, Zhangyang Wang, Konstantinos N Plataniotis, Yao Zhao, and Yunchao Wei. Diffusion4d: Fast spatial-temporal consistent 4d generation via video diffusion models. *arXiv preprint arXiv:2405.16645*, 2024.
- Isabella Liu, Hao Su, and Xiaolong Wang. Dynamic gaussians mesh: Consistent mesh reconstruction from dynamic scenes. *arXiv preprint arXiv:2404.12379*, 2024.
- Cheng-You Lu, Peisen Zhou, Angela Xing, Chandradeep Pokhariya, Arnab Dey, Ishaan Nikhil Shah, Rugved Mavidipalli, Dylan Hu, Andrew I Comport, Kefan Chen, et al. Diva-360: The dynamic visual dataset for immersive neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22466–22476, 2024.
- Wei Mao, Richard Hartley, Mathieu Salzmann, et al. Neural sdf flow for 3d reconstruction of dynamic scenes. In *The Twelfth International Conference on Learning Representations*, 2024.
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8280–8290, June 2022.
- Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4d reconstruction by learning particle dynamics. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5379–5389, 2019.
- Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5865–5874, 2021a.
- Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228*, 2021b.

- Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10318–10327, 2021.
- Saskia Rabich, Patrick Stotko, and Reinhard Klein. Fpo++: efficient encoding and rendering of dynamic neural radiance fields by analyzing and enhancing fourier plenotrees. *The Visual Computer*, 40(7):4777–4788, 2024.
- Jiawei Ren, Cheng Xie, Ashkan Mirzaei, Karsten Kreis, Ziwei Liu, Antonio Torralba, Sanja Fidler, Seung Wook Kim, Huan Ling, et al. L4gm: Large 4d gaussian reconstruction model. *Advances in Neural Information Processing Systems*, 37:56828–56858, 2024.
- Lu Sang, Zehranaz Canfes, Dongliang Cao, Riccardo Marin, Florian Bernard, and Daniel Cremers. 4deform: Neural surface deformation for robust shape interpolation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 6542–6551, 2025.
- Tianchang Shen, Jun Gao, Kangxue Yin, Ming-Yu Liu, and Sanja Fidler. Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.
- Tianchang Shen, Jacob Munkberg, Jon Hasselgren, Kangxue Yin, Zian Wang, Wenzheng Chen, Zan Gojcic, Sanja Fidler, Nicholas Sharp, and Jun Gao. Flexible isosurface extraction for gradient-based mesh optimization. *ACM Trans. Graph.*, 42(4), jul 2023. ISSN 0730-0301. doi: 10.1145/3592430. URL <https://doi.org/10.1145/3592430>.
- Lixin Shi, Haitham Hassanieh, Abe Davis, Dina Katabi, and Fredo Durand. Light field reconstruction using sparsity in the continuous fourier domain. *ACM Transactions on Graphics (TOG)*, 34(1):1–13, 2014.
- Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.
- Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5481–5490. IEEE, 2022.
- Hengyi Wang, Jingwen Wang, and Lourdes Agapito. Morpheus: Neural dynamic 360deg surface reconstruction from monocular rgb-d video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20965–20976, 2024.
- Liao Wang, Jiakai Zhang, Xinhang Liu, Fuqiang Zhao, Yanshun Zhang, Yingliang Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Fourier plenotrees for dynamic radiance field rendering in real-time. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13524–13534, 2022.
- Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Proceedings of the 35th International Conference on Neural Information Processing Systems, NIPS ’21*, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN 9781713845393.
- Yifan Wang, Peishan Yang, Zhen Xu, Jiaming Sun, Zhanhua Zhang, Yong Chen, Hujun Bao, Sida Peng, and Xiaowei Zhou. Freetimegs: Free gaussian primitives at anytime anywhere for dynamic scene reconstruction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 21750–21760, 2025.
- Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3295–3306, 2023.



- Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20310–20320, 2024.
- Rundi Wu, Ruiqi Gao, Ben Poole, Alex Trevithick, Changxi Zheng, Jonathan T Barron, and Aleksander Holynski. Cat4d: Create anything in 4d with multi-view video diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 26057–26068, 2025.
- Jiawei Xu, Zexin Fan, Jian Yang, and Jin Xie. Grid4d: 4d decomposed hash encoding for high-fidelity dynamic gaussian splatting. *Advances in Neural Information Processing Systems*, 37: 123787–123811, 2024.
- Hao Yan, Zhihui Ke, Xiaobo Zhou, Tie Qiu, Xidong Shi, and Dadong Jiang. Ds-nerv: Implicit neural video representation with decomposed static and dynamic codes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23019–23029, 2024.
- Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 20331–20341, 2024.
- Yuxin Yao, Siyu Ren, Junhui Hou, Zhi Deng, Juyong Zhang, and Wenping Wang. Dynosurf: Neural deformation-based temporally consistent dynamic surface reconstruction. In *European Conference on Computer Vision*, pp. 271–288. Springer, 2024.
- Kai Ye, Chong Gao, Guanbin Li, Wenzheng Chen, and Baoquan Chen. Geosplatting: Towards geometry guided gaussian splatting for physically-based inverse rendering. *arXiv preprint arXiv:2410.24204*, 2024.
- Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.
- Shuai Zhang, Guanjun Wu, Xinggang Wang, Bin Feng, and Wenyu Liu. Dynamic 2d gaussians: Geometrically accurate radiance fields for dynamic objects, 2024. URL <https://arxiv.org/abs/2409.14072>.
- Chengwei Zheng, Lixin Xue, Juan Zarate, and Jie Song. Gaustar: Gaussian surface tracking and reconstruction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 16543–16553, 2025.
- Ziyue Zhu, Shenlong Wang, Jin Xie, Jiang-jiang Liu, Jingdong Wang, and Jian Yang. Voxelsplat: Dynamic gaussian splatting as an effective loss for occupancy and flow prediction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 6761–6771, 2025.