

Parametric SDF for Dynamic Surface Reconstruction

Chong Gao^{1*}, Kai Ye^{2*}, Qiyu Dai², Yiming Shao³, Qiong Zeng³, Ding Liang⁴,
Yan-Pei Cao⁴, Guanbin Li^{1,5†}, and Wenzheng Chen^{6,7†}

¹Sun Yat-sen University ²Peking University ³Shandong University ⁴VAST
⁵Shenzhen Loop Area Institute ⁶Wangxuan Institute of Computer Technology,
Peking University ⁷Beijing Academy of Artificial Intelligence, Beijing, China
gaoch63@mail2.sysu.edu.cn, wenzhengchen@pku.edu.cn

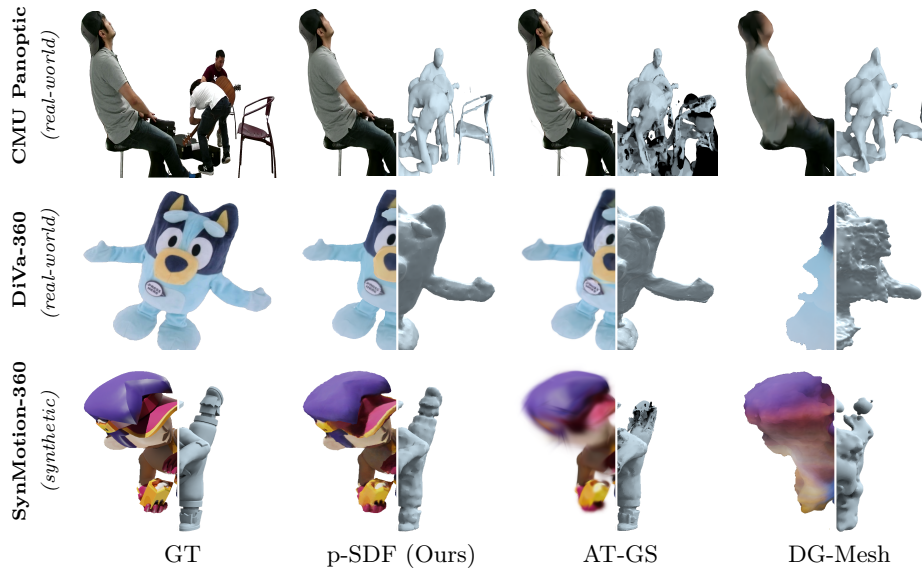


Fig. 1: Qualitative Comparison of Dynamic Surface Reconstructions. Compared with previous works, p-SDF recovers more faithful geometry and appearance, especially in scenarios with large non-rigid deformations and complex topological changes.

Abstract. Reconstructing high-fidelity surfaces of dynamic scenes remains a critical challenge. While current methods excel at novel view synthesis, they often struggle to recover accurate and stable geometry, particularly under large non-rigid deformations. This results in noisy meshes that are unsuitable for downstream 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 where each curve models a temporally evolving SDF trajectory. This parametric SDF representation provides a principled way to capture complex temporal variations, naturally enforcing smoothness and continuity in shape dynamics.

* Equal contribution.

† Corresponding authors.

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 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 lighting and occlusions. Experiments on both synthetic and real-world scenes demonstrate that our method achieves state-of-the-art geometric accuracy. The project page for p-SDF is available at <https://timchonggao.github.io/psdf>.

Keywords: Dynamic Surface Reconstruction · Parametric SDF · Differentiable Rendering

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. Achieving this remains highly challenging, as dynamic scenes often involve complex non-rigid motions, large-scale dynamics, and intricate topological changes.

Recent advances in scene representations, such as Neural Radiance Fields [23] and 3D Gaussian Splatting [13], have dramatically advanced novel view synthesis (NVS). Their dynamic extensions [1, 7, 9, 26, 38] further enable photorealistic NVS across time. Nevertheless, these methods are designed primarily for appearance modeling, often producing noisy or temporally unstable underlying geometry. Two lines of work have explored ways to improve the quality of dynamic surface reconstruction. The first strategy adopts a Lagrangian perspective, tracking 3D trajectories through deformation fields. This involves mapping a canonical template to the dynamic scene via global deformations [19, 42] or incremental deformations [2, 35, 43]. By explicitly disentangling appearance from motion, this approach maintains temporal consistency across frames. However, it is inherently limited by the fixed-topology template and the smoothing effect of deformation MLPs, making it difficult to handle topological changes and highly complex non-rigid dynamics. To bypass explicit tracking, the second strategy adopts an Eulerian perspective by factorizing the 4D spatio-temporal volume into plane or tensor features [9, 28]. Without explicit motion modeling, the representation is unconstrained by continuous motion functions, enabling it to handle topological changes or discontinuous events. However, without temporal regularization, mesh quality often becomes unstable under complex dynamics.

In this work, we propose parametric Signed Distance Function(p-SDF), a novel dynamic scene representation that reconstructs high-fidelity surfaces across time. Our key intuition is that while 3D shapes deform, the evolution of the Signed Distance Function (SDF) at any fixed spatial coordinate follows a continuous 1D trajectory (see Fig. 2; refer to Appendix Sec. A for a more detailed

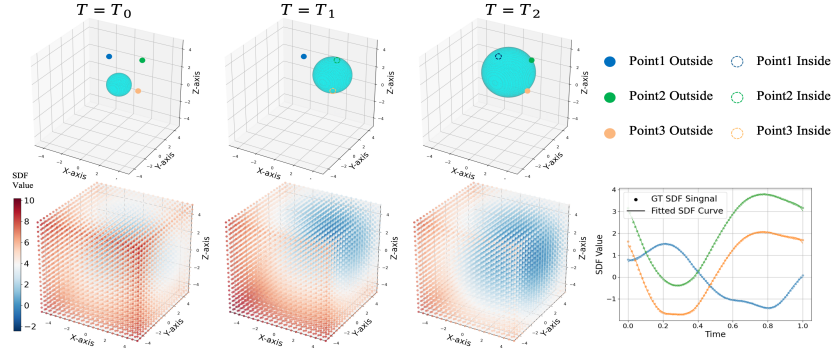


Fig. 2: Continuous SDF Trajectory. As a dynamic scene evolves, the SDF value at any fixed spatial coordinate follows a continuous temporal trajectory. In the example of a moving sphere (left), spatial points transition between positive (outside) and negative (inside) SDF values, forming distinct curves. We optimize p-SDF with a compact set of basis functions (Sec. 3.3) to represent these curves. As shown in the bottom-right plot, p-SDF closely follows the GT SDF trajectories.

analysis). We thus generalize static SDF fields into time-dependent parametric curves. Unlike Lagrangian methods that struggle with learning complex 3D deformation fields, p-SDF reformulates dynamic surface reconstruction as a 1D spatio-temporal signal fitting problem. p-SDF uses compact basis functions to represent these curves, naturally enforcing temporal coherence while handling large deformations and topological changes.

Specifically, we represent dynamic geometry using a grid where each vertex stores basis function parameters: low-order polynomials for global deformations and Fourier bases for high-frequency details. At any timestamp, a static SDF field can be queried to extract an explicit mesh via differentiable iso-surfacing [30]. Separately, an efficient temporal tri-plane appearance field models the scene’s appearance by predicting PBR material properties. The resulting mesh and materials are then fed into a differentiable renderer for end-to-end optimization against 2D images. The explicit disentanglement of geometry, materials, and lighting constrains the network to reconstruct physically accurate surfaces.

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

2 Related Work

2.1 Neural 3D Representations

Recent breakthroughs in Neural Radiance Fields [23] and 3D Gaussian Splatting [13] have revolutionized static novel-view synthesis. However, their primary focus on appearance modeling often comes at the expense of geometric accuracy. NeRF-based methods leverage coordinate-based networks to represent volume

density and color. To extract usable geometry, several works incorporate signed distance functions (SDFs) as underlying representations [34, 35, 40], enabling surface recovery via Marching Cubes [20]. Gaussian-based methods represent scenes as explicit 3D primitives. To bridge the gap between discrete points and continuous surfaces, techniques such as surface-alignment regularization [4, 5, 8, 10], Poisson reconstruction [5, 8], and SDF fusion [3, 11] have been introduced. However, existing extraction pipelines, such as Poisson reconstruction or SDF fusion, typically fail to maintain stable topology or a fixed vertex count across dynamic frames. Despite these advancements in static reconstruction, extending these representations to dynamic scenarios remains a formidable challenge.

2.2 Dynamic Surface Reconstruction

The central challenge in dynamic surface reconstruction is modeling temporal evolution while preserving spatial fidelity. Existing approaches generally follow three paradigms: Global Deformation [19, 25, 26, 37, 38, 42], Incremental Optimization [2, 35, 43], and 4D Space-Time Volumes [1, 9, 22, 28].

The global deformation paradigm establishes a mapping between a static canonical space and dynamic observations at time t . Many methods in this category are designed for monocular settings where each timestamp provides a single image from a randomly sampled viewpoint. For instance, DG-Mesh [19] deforms canonical 3D Gaussians into target frames and recovers surfaces via per-frame Poisson reconstruction. Dynamic 2DGS [42] refines this by employing surface-aligned 2D surfels with TSDF fusion. However, these methods are fundamentally limited by the fixed topology of their canonical templates. Furthermore, their deformations rely on continuous MLPs that can oversmooth temporal motion. We adapt these baselines to the multi-view setting for fair comparison.

Instead of a global mapping, incremental optimization reconstructs scenes frame-by-frame, focusing on deformations between adjacent timestamps. Approaches like NeuS2 [35] and AT-GS [2] typically propagate weights from the previous frame to the current one. While capable of handling large motions and topological changes, these methods are constrained by their reliance on auxiliary supervision, such as optical flow [43], to curb error accumulation.

The 4D space-time volume paradigm decomposes the 4D domain into efficient spatio-temporal representations. Methods like Tensor4D [28] factorize 4D volumes into planes to represent density and color at any (x, y, z, t) . As a highly flexible paradigm, it bypasses explicit motion modeling and fixed templates, enabling the recovery of complex dynamics and topological changes. However, lacking strong structural priors, 4D volumes suffer from temporal flickering.

Similar to our work, SDFFlow [22] models SDF deformations but relies on an integral formulation that is computationally expensive and prone to error accumulation. This makes it unsuitable for long sequences and high-fidelity reconstruction. In contrast, our method directly parameterizes the SDF trajectories with basis functions, inherently enforcing continuity regularization. The static SDF at a given timestamp is efficiently obtained by querying the function, avoiding costly explicit integration that accumulates errors.

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 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 discuss the implementation details in Sec. 3.5.

3.1 Background

Since our method integrates explicit surface reconstruction with basis-function-based temporal modeling, we first introduce preliminaries from both domains.

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 Ω 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 [24] define SDF values on a 3D grid and use differentiable iso-surfacing [29, 30] to extract a triangle mesh. The resulting mesh can then be rendered via a differentiable rasterizer [15]. While highly effective for static scenes, naively extending such pipelines to dynamic scenes, *i.e.*, applying a per-frame reconstruction strategy, is computationally expensive.

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 pre-defined temporal basis functions and $\{w_k\}_{k=1}^K$ are learnable coefficients. With appropriate choices of basis functions, a small set of coefficients compactly encodes complex temporal behavior while naturally encouraging smoothness.

Common choices for the basis include polynomial and Fourier functions. Specifically, the polynomial basis employs a combination of power functions:

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

which captures smooth and low-frequency trends. Alternatively, Fourier basis functions use periodic functions based on sin and cos:

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

where ω_k is the predefined frequency for the k -th periodic function, and a_k and b_k are the coefficients. The Fourier basis is well-suited for periodic or high-frequency variations, as shown in the literature [16, 27, 31–33]. This decomposition improves interpretability: polynomial coefficients capture static and low-order SDF trends, while Fourier coefficients capture low- and high-frequency temporal variations.

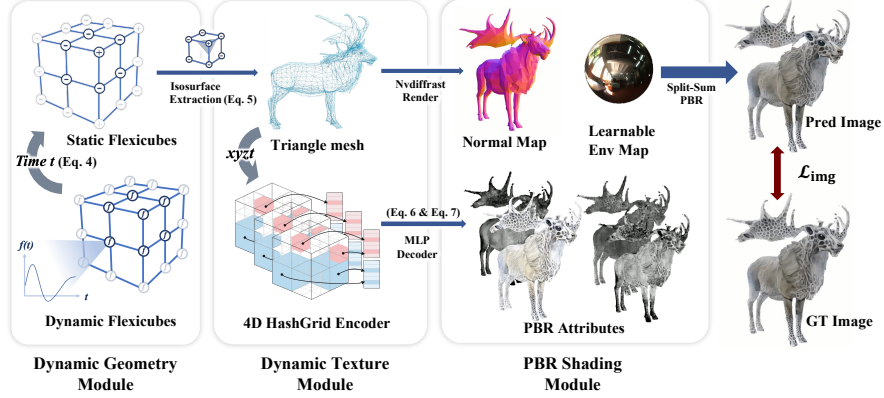


Fig. 3: Pipeline. For a timestamp t , p-SDF queries the parametric function at each grid node to obtain the SDF value (Eq. (4)). We use FlexiCubes [30] to extract the mesh $M_t = (V_t, F_t)$ (Eq. (5)) and query the Appearance Field (Sec. 3.4) to obtain (c, m, r, o) (Eq. (7)) at a point $(\mathbf{x}, t)$. A differentiable PBR shader renders multi-view images, with gradients propagating through all modules end-to-end (Sec. 3.5).

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 motion 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. This parameterization naturally enforces temporal coherence in the SDF evolution while retaining flexibility for large deformations and topological changes.

Pipeline. Fig. 3 illustrates the full system. Given a timestamp t , we first obtain per-node SDF values by evaluating basis functions and then extract an explicit mesh using differentiable iso-surfacing [30]. Concurrently, to model appearance, we apply a temporal tri-plane appearance field and query PBR attributes at each mesh vertex. 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).

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 $N \times N \times N$ nodes $\{\mathbf{p}_i\}_{i=1}^{N^3}$. Unlike static methods that assign a single SDF value to each node, we model the SDF value at node $\mathbf{p}_i$ as a continuous function of time: $s(\mathbf{p}_i, t) : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}$. Following prior work on temporal signal modeling [18], we adopt a hybrid basis combining polynomials and Fourier terms to represent $s(\mathbf{p}_i, t)$:

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

where N_p is the number of polynomial terms, and N_f is the number of Fourier components. 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 node coefficients.

Mesh Extraction. Given our time-varying SDF grid and a timestamp t , we evaluate the basis functions to obtain per-node SDF values. A surface mesh is then extracted using FlexiCubes, a differentiable isosurface extraction method [30]:

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

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

Discussion. This parametric surface-mesh design provides several important benefits: **(i)** the basis formulation inherently enforces temporal coherence, while the combination of polynomial and Fourier bases captures both large-scale deformations and fine-grained high-frequency motions, offering robustness in complex dynamic scenes; **(ii)** we model the geometry by storing per-node 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 AT-GS [2].

3.4 Appearance Representation

Formulation. Having established a continuous representation for dynamic geometry, we now model the time-varying appearance and material properties. 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 because of lighting variation, occlusion, and surface reflectance effects. 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 Φ_{app} into physically based material properties:

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

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

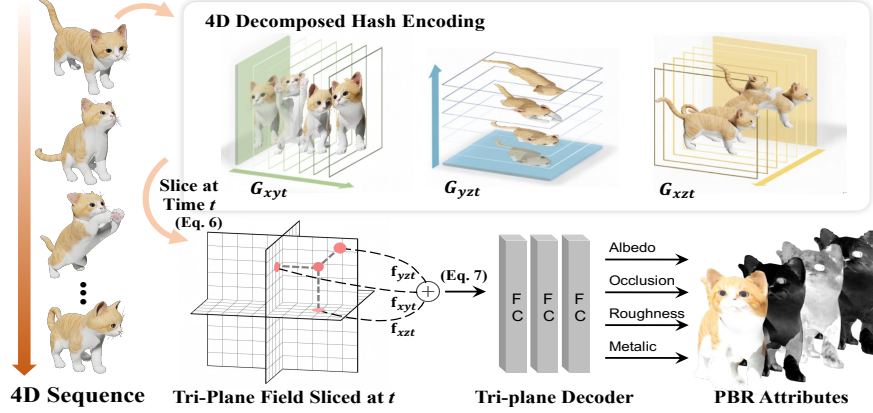


Fig. 4: Temporal Tri-plane Appearance Field. Left: an input multi-view video. Middle: three multi-resolution spatio-temporal hash grids $G_{xyt}, G_{xzt}, G_{yzt}$ that are sliced at time t to produce an instantaneous tri-plane. Right: features sampled from the three planes are concatenated and fed into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

Rendering. At time t , we rasterize the mesh $M_t = (V_t, F_t)$ extracted by the geometry module (Sec. 3.3) and generate a G-buffer of world-space positions $\mathbf{x}$ and surface normals $\mathbf{n}$. For each visible surface point, the Temporal Tri-plane Appearance Field is queried at $(\mathbf{x}, t)$ to produce material properties $(\mathbf{c}, m, r, o)$ (Eq. (7)). These attributes, combined with the mesh geometry and a shared learnable environment map, are passed to a differentiable PBR shader, following Nvdiffrac [24], to compute the final rendered RGB image. This design allows end-to-end optimization of both geometry and appearance against multi-view 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 compared with directly modeling 4D appearance; (ii) with precise geometry, this simple representation can disentangle shape and appearance and produce high-quality material and lighting estimates.

3.5 Implementation Details

Optimization Pipeline. The input to p-SDF consists of multi-view RGB videos, foreground masks, and calibrated camera intrinsics/extrinsics. At each iteration, we query p-SDF at time t , extract a mesh with differentiable FlexiCubes [30], rasterize it with nvdiffrast [15], query PBR attributes at visible $(\mathbf{x}, t)$ surface points, and render the target view with a differentiable PBR shader following Nvdiffrac [24]. Gradients from the RGB and mask losses, together with regularizers, are backpropagated end-to-end to the geometry, appearance, and lighting parameters.

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 Sec. B.

Model Details. We represent p-SDF with a 96^3 grid, 6 polynomial basis functions, and a set of Fourier components whose number is chosen per scene from the range $[18, 100]$ based on motion complexity. For the Temporal Tri-plane Appearance Field, we adapt the multi-resolution hash grids following Grid4D [37], 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 [39] 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 [14] with exponential decay.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate our method on three diverse datasets to demonstrate its robustness across synthetic and real-world scenarios with complex dynamics:

SynMotion-360. We first curate a *new synthetic benchmark* designed to evaluate dynamic surface reconstruction under large, non-rigid deformations that existing datasets rarely capture. Specifically, we select 7 animated assets from Objaverse [6, 17], each comprising 70–300 frames rendered via Blender from 38 viewpoints. For each sequence, 32 views are used for training and 6 for testing. Additional dataset visualizations are provided in Appendix Sec. J.

DiVa-360. This is a real-world multi-view dataset featuring dynamic tabletop scenes captured by synchronized RGB cameras [21]. While it provides synchronized videos and foreground masks, ground-truth meshes are unavailable. We evaluate our method on 4 representative sequences, using 31 cameras for training and 6 held-out views for evaluation.

CMU Panoptic Studio. This widely used real-world dataset captures human dynamics [12]. Following the protocol in SDFFlow [22], we focus on challenging multi-person interactions. Each sequence consists of 24 frames recorded from 10 calibrated cameras, and all views are used for training.

Metrics. We evaluate performance from two aspects: *geometric reconstruction accuracy* and *novel-view synthesis quality*. Following [30], geometric quality is quantified using Chamfer Distance (CD) and F-Score (F_1) for spatial alignment, while Edge Chamfer Distance (ECD) and Edge F-Score (EF_1) assess boundary fidelity and topological connectivity. To assess mesh structural integrity, we report the percentage of sliver triangles using Aspect Ratio (AR) and Minimum Angle (MA) criteria: the AR metric counts triangles with aspect ratio above 4.0,

and the MA metric counts triangles with minimum angle below 10° . Regarding synthesis quality, we report $PSNR$, $SSIM$ [36], and $LPIPS$ [41] to capture pixel-level accuracy, structural similarity, and perceptual realism, respectively. Held-out test images are used only for novel-view synthesis evaluation; geometry metrics are computed from ground-truth meshes or point clouds when available, and from mesh-quality statistics otherwise.

Baselines. We evaluate our method against state-of-the-art (SOTA) baselines from two predominant paradigms: *Global Deformation* and *Incremental Optimization*. For the former, we compare with Dynamic 2DGS [42], which uses surface-aligned 2D surfels with TSDF fusion, and DG-Mesh [19], which performs per-frame Poisson reconstruction from deformed Gaussians. For the latter, we include NeuS2 [35], a representative SDF-based method using per-frame hash grids, and AT-GS [2], which employs neural transformation caches for frame-by-frame propagation. While SDFFlow [22] is conceptually related to our trajectory modeling, it is excluded due to its prohibitive optimization costs and error accumulation as discussed in Sec. 2. GauSTAR [43] requires additional depth supervision, and DynaSurfGS is not included because its code was unavailable at the time of evaluation. All reported baselines are trained in the same multi-view video setting. For monocular baselines [19,42], we extend their data loaders to aggregate supervision from all training cameras at each timestamp; the main tables report these multi-view-adapted results. A controlled monocular-vs.-multi-view comparison is provided in Appendix Sec. F. This diverse selection of baselines supports a rigorous assessment of our model’s capacity to handle complex dynamics and high-fidelity dynamic mesh reconstruction.

4.2 Experimental Results

Quantitative Results. We provide comprehensive quantitative comparisons with all baselines in Tabs. 1 and 2, with per-scene details provided in Appendix Tables A1–A3.

Geometric Accuracy. Tab. 1 summarizes the quantitative geometric evaluation across three diverse datasets. We employ a tiered metric system based on ground-truth (GT) availability: mesh-to-mesh metrics (CD , F_1 , ECD , EF_1) for SynMotion-360; point-to-mesh (CD , F_1) and quality metrics (AR, MA) for CMU Panoptic Studio; and mesh quality (AR, MA) for the real-world DiVa-360. Our p-SDF consistently achieves state-of-the-art performance. Unlike incremental baselines (NeuS2, AT-GS) limited by frame-wise propagation or template-based methods (DG-Mesh, D-2DGS) restricted by fixed topologies, our parametric modeling offers the flexibility to represent non-linear trajectories and complex topological changes. Furthermore, while incremental methods suffer from accumulation error under sparse supervision (CMU Panoptic), our global curve modeling ensures temporal coherence and consistency. Finally, leveraging end-to-end supervision via differentiable FlexiCubes allows our framework to avoid floaters and degenerate triangles common in Gaussian-based extractions. The large AR/MA gap is expected: these values should be interpreted as mesh-quality

diagnostics rather than direct reconstruction-accuracy metrics. p-SDF optimizes the SDF field and FlexiCubes extraction parameters end-to-end, whereas baseline meshes rely on post-hoc extraction pipelines such as Poisson reconstruction, TSDF fusion, or isosurfacing, which are more prone to sliver triangles.

Table 1: Quantitative Evaluation of Geometry Quality. We evaluate geometric reconstruction across three diverse datasets. Our method achieves the strongest overall performance in reconstruction accuracy and mesh structural integrity. Detailed metric definitions are available in Sec. 4.1.

Method	SynMotion-360				CMU Panoptic Studio				DiVa-360	
	CD↓	F1↑	ECD↓	EF1↑	CD↓	F1↑	AR↓	MA↓	AR↓	MA↓
Dynamic-2DGS	2.104	0.754	3.802	0.494	2.000	0.762	22.656	12.435	23.156	12.830
DG-Mesh	2.957	0.661	4.740	0.438	<u>1.041</u>	<u>0.788</u>	22.508	12.894	22.829	13.207
NeuS2	<u>1.203</u>	0.908	3.220	0.593	7.654	0.729	21.631	12.563	22.530	12.976
AT-GS	1.602	0.882	3.227	0.566	6.525	0.687	<u>19.359</u>	<u>11.353</u>	<u>19.234</u>	<u>11.206</u>
Ours	1.157	<u>0.893</u>	3.170	0.596	0.026	0.896	4.983	0.678	3.126	0.396

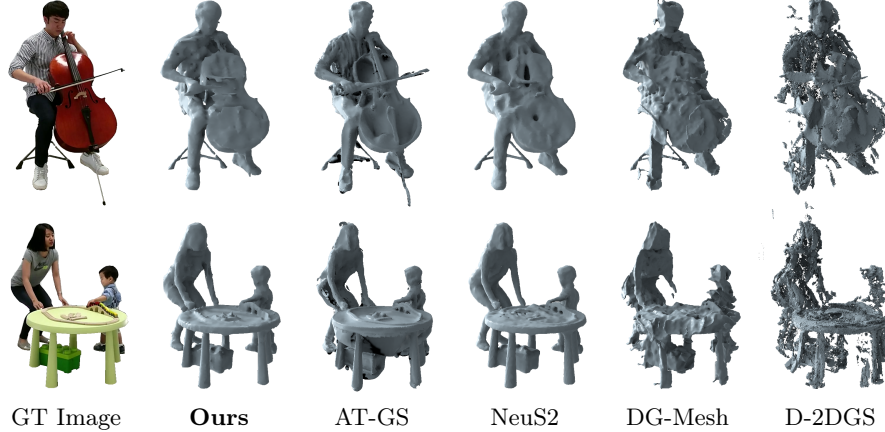


Fig. 5: Qualitative comparison on the *CMU Panoptic Studio* dataset. We evaluate geometric reconstruction performance across challenging scenes featuring complex human interactions. Our method consistently recovers accurate geometry, whereas existing surface reconstruction baselines suffer from noticeable artifacts.

Rendering Quality. For novel view synthesis (Tab. 2), we report PSNR, SSIM, and LPIPS on SynMotion-360 and DiVa-360. CMU Panoptic Studio is excluded from this evaluation as all available views are consumed for training. On SynMotion-360, our *p*-SDF achieves competitive results, ranking second in PSNR and SSIM while securing the best LPIPS. The high photometric scores of AT-GS can be attributed to its adaptive-density Gaussian primitives, which are highly optimized for capturing fine-grained texture details. However, as indicated by its suboptimal geometric metrics in Tab. 1, such rendering-centric representations often compromise surface fidelity for photometric overfitting. In contrast, our *p*-SDF offers a more principled trade-off, delivering top-tier perceptual realism alongside far superior geometric accuracy. On the real-world DiVa-360 dataset, our framework establishes new state-of-the-art performance across all

rendering metrics. This superiority in real-world tabletop scenarios highlights the advantage of our PBR-aligned appearance modeling. By leveraging precise geometry as a prior, p-SDF effectively resolves appearance ambiguities and surface artifacts, ensuring robust generalization across diverse environments.

Table 2: Quantitative Evaluation of Rendering Quality. Compared to prior works, p-SDF achieves state-of-the-art performance across all metrics on the real-world DiVa-360 dataset and remains highly competitive on synthetic sequences.

Method	SynMotion-360			DiVa-360		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Dynamic-2DGS	29.489	0.959	0.062	17.513	0.910	0.187
DG-Mesh	25.719	0.936	0.081	17.734	0.919	0.183
NeuS2	30.070	0.967	0.030	26.087	0.950	0.056
AT-GS	32.671	0.977	0.045	<u>27.418</u>	<u>0.956</u>	0.067
Ours	<u>32.153</u>	<u>0.974</u>	0.028	27.584	0.957	0.049

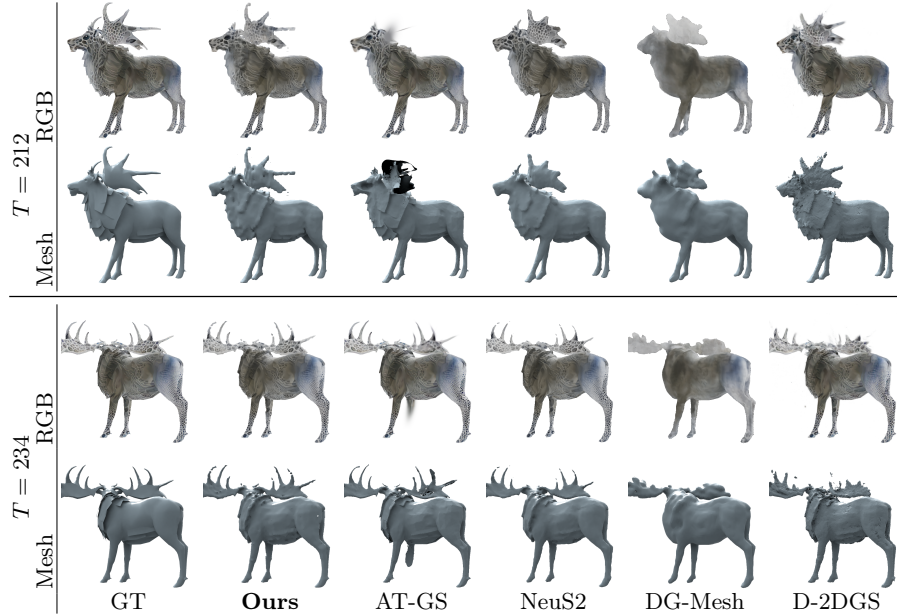


Fig. 6: Qualitative comparison of visual and geometry quality on the *SynMotion-360* sequence. We show rendered images and the corresponding reconstructed meshes at two different time steps ($T = 212, 234$). Compared to previous methods, our approach accurately recovers fine-grained geometric details.

Qualitative Results. We provide qualitative comparisons across three datasets in Figs. 1 and 5 to 7, with additional results in Appendix Figs. A3–A6.

Robustness to Rapid Motion. As illustrated by the *SynMotion-360* Lego sequence in Fig. 1, where motions are highly nonlinear and fast, our method demonstrates robustness. Incremental baselines like AT-GS suffer from severe

motion blur in fast-moving regions, while template-based methods such as DG-Mesh fail to track large deformations, resulting in fractured or missing meshes. In contrast, our p-SDF leverages parametric curves to fit the SDF trajectories, enabling it to precisely capture rapidly evolving geometric boundaries and reconstruct complete dynamic surfaces.

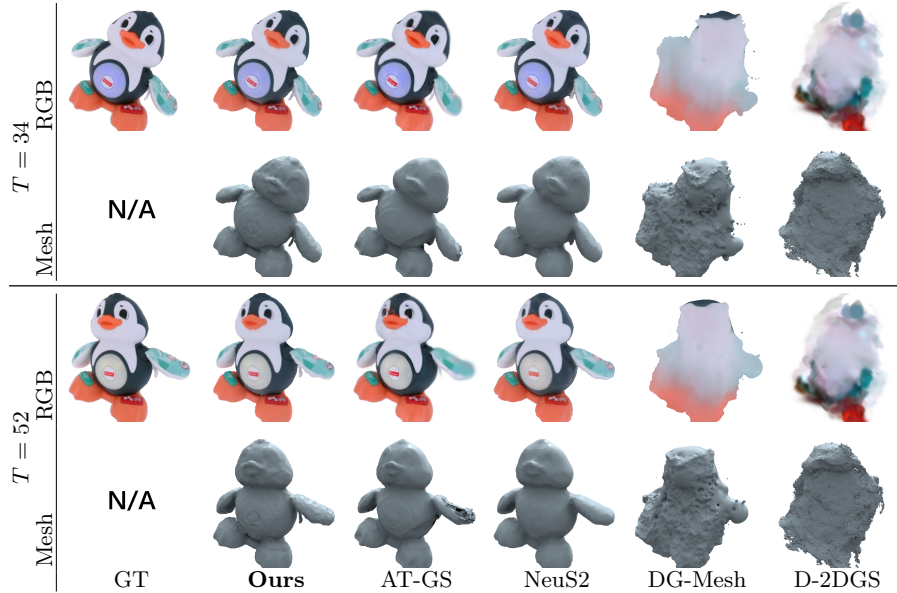


Fig. 7: Qualitative comparison of visual and geometry quality on the *DiVa-360* sequence. We show rendered images and the corresponding reconstructed meshes at two different time steps ($T = 34, 52$). Compared to previous methods, our approach (p-SDF) achieves more realistic appearance and cleaner geometry with fewer artifacts.

High-Fidelity Geometry. A key strength of our approach is the ability to recover fine-grained structural details and avoid artifacts. In the CMU Panoptic sequence (Fig. 1), p-SDF is the only method in this comparison that reconstructs thin structures like the chair legs, whereas baselines produce noisy or over-smoothed results. Our method preserves structural integrity for sharp features, such as the antlers in the SynMotion-360 deer sequence (Fig. 6). On the DiVa-360 sequence (Fig. 7), p-SDF recovers tiny details such as the logo texture under the penguin’s arm, which are lost or distorted by other methods.

4.3 Further Analysis

Ablations. We analyze basis composition, grid resolution, and Fourier bandwidth in Appendix Tables A4–A6. Fourier bases are critical for dynamic geometry: the polynomial-only variant reaches only 25.91 dB PSNR, Fourier-only improves to 35.90 dB, and the full mixed basis reaches 36.34 dB. For spatial discretization, 96^3 provides our default quality-cost trade-off between the coarse 64^3 grid and

the costlier 128^3 grid. For temporal bandwidth, smooth-motion Toy saturates at lower N_f , whereas fast-motion Lego benefits from larger bandwidth.

Computational Cost. We report training cost, peak VRAM, and inference speed in Appendix Tables A9 and A10. p-SDF uses more VRAM (19.7 GB) because its dense parametric grid stores coefficients for the full temporal SDF field, but its training time remains practical at 121 min on an RTX 4090. At inference, p-SDF runs at 11.4 FPS, slower than Gaussian baselines due to mesh extraction and PBR material evaluation, but faster than NeuS2 (8.6 FPS).

Topology Changes. We provide topology-change evidence in Appendix Figs. A1 and A2. In a torus-to-sphere sequence, p-SDF fits the sharp SDF sign transition caused by hole closure and reconstructs the topology change from multi-view observations. This behavior follows from modeling per-location SDF trajectories directly, without relying on a fixed-connectivity canonical template.

Large Motion and Temporal Smoothness. Appendix Sec. E evaluates large-motion robustness and temporal interpolation. On the fast-motion Lego sequence, the motion-velocity-versus-CD analysis shows that p-SDF maintains lower geometric error in high-speed intervals and reconstructs more complete surfaces. Under an even-frame training and odd-frame testing protocol, p-SDF also achieves the lowest interpolation CD on both smooth Toy and fast-motion Lego, supporting the temporal consistency of the learned SDF trajectories.

Monocular Setting. Appendix Sec. H further evaluates p-SDF under the monocular protocol. A qualitative comparison with representative monocular baselines shows that p-SDF remains competitive for smooth dynamics, while additional analyses reveal noisy surfaces and degradation under large motion. These results clarify that the temporal SDF prior helps under weak supervision, but calibrated multi-view consistency remains essential for high-fidelity geometry.

Additional Results. Appendix Sec. I provides additional qualitative Tensor4D comparisons and NHR dataset results, complementing our evaluation with a 4D volume baseline and human-centric multi-view captures.

4.4 Relighting

To further validate the accuracy and reliability of our dynamic reconstruction, we conduct relighting experiments using our reconstructed geometry and PBR materials. As shown in Fig. 8, the reconstructed results enable high-quality relighting for dynamic scenes.

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 temporal coherence while remaining flexible to large deformations and topological changes.

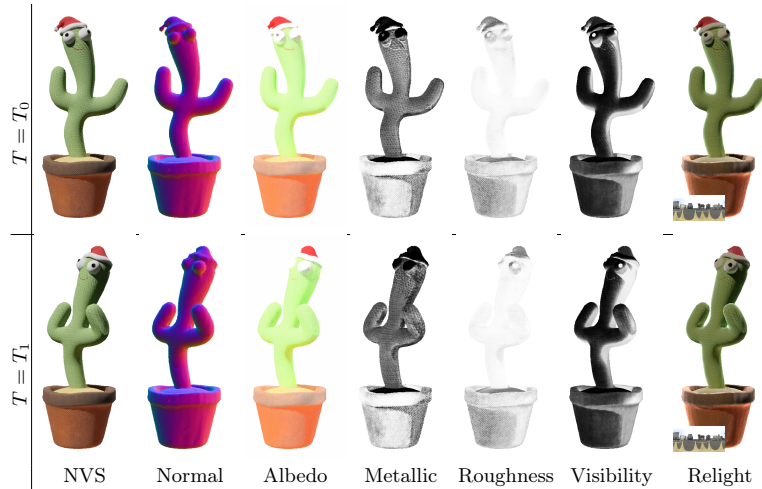


Fig. 8: PBR material reconstruction and relighting results. Our method reconstructs plausible PBR materials (albedo, metallic and roughness) from multi-view videos, supporting re-lighting applications with new environment maps.

Combined with physically based appearance modeling, our framework yields accurate geometry and robust material decomposition in highly dynamic conditions, and supports downstream applications such as relighting. Experiments on synthetic and real-world calibrated multi-view videos show that p-SDF establishes state-of-the-art geometric accuracy while maintaining competitive novel-view synthesis quality. We believe p-SDF offers a promising direction for dynamic surface reconstruction and its applications.

Limitations. Our method relies on multi-view consistency. In extremely sparse scenarios (e.g., monocular data), reconstruction accuracy inevitably degrades, although the analytical basis functions of p-SDF provide a smooth prior that prevents global structural collapse. Detailed analysis of this monocular setting is provided in Appendix Sec. H. As an Eulerian representation, p-SDF models the temporal evolution of SDF values at stationary grid nodes; thus, it does not inherently provide point-to-point correspondences, such as vertex tracking, across frames. Nevertheless, the temporal continuity of our basis functions naturally supports smooth temporal interpolation and slow-motion synthesis. These properties and their implications are analyzed in Appendix Sec. E.

Acknowledgements

This work is supported in part by the National Key R&D Program of China (Nos. 2024YFB3908503 and 2024YFB3908500), in part by the National Natural Science Foundation of China (Nos. 62322608 and 62595772), in part by the Beijing Science and Technology Program (No. Z251100008125028), and in part by the Shenzhen Loop Area Institute under grant FPF10120260001.

References

1. Cao, A., Johnson, J.: Hexplane: A fast representation for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 130–141 (2023)
2. Chen, D., Oberson, B., Feldmann, I., Schreer, O., Hilsmann, A., Eisert, P.: 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)
3. Chen, H., Wei, F., Li, C., Huang, T., Wang, Y., Lee, G.H.: Vcr-gaus: View consistent depth-normal regularizer for gaussian surface reconstruction. *Advances in Neural Information Processing Systems* **37**, 139725–139750 (2024)
4. Choi, J., Lee, Y., Lee, H., Kwon, H., Manocha, D.: Meshgs: Adaptive mesh-aligned gaussian splatting for high-quality rendering. In: Proceedings of the Asian Conference on Computer Vision. pp. 3310–3326 (2024)
5. Dai, P., Xu, J., Xie, W., Liu, X., Wang, H., Xu, W.: High-quality surface reconstruction using gaussian surfels. In: ACM SIGGRAPH 2024 conference papers. pp. 1–11 (2024)
6. Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., VanderBilt, E., Schmidt, L., Ehsani, K., Kembhavi, A., Farhadi, A.: Objaverse: A universe of annotated 3d objects. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13142–13153 (2023)
7. Duan, Y., Wei, F., Dai, Q., He, Y., Chen, W., Chen, B.: 4d-rotor gaussian splatting: Towards efficient novel-view synthesis for dynamic scenes. In: Proc. SIGGRAPH (July 2024)
8. Fan, L., Yang, Y., Li, M., Li, H., Zhang, Z.: Trim 3d gaussian splatting for accurate geometry representation. *arXiv preprint arXiv:2406.07499* (2024)
9. Fridovich-Keil, S., Meanti, G., Warburg, F.R., Recht, B., Kanazawa, A.: 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)
10. Guédon, A., Lepetit, V.: Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5354–5363 (2024)
11. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: ACM SIGGRAPH 2024 conference papers. pp. 1–11 (2024)
12. Joo, H., Liu, H., Tan, L., Gui, L., Nabbe, B., Matthews, I., Kanade, T., Nobuhara, S., Sheikh, Y.: Panoptic studio: A massively multiview system for social motion capture. In: Proceedings of the IEEE international conference on computer vision. pp. 3334–3342 (2015)
13. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139–1 (2023)
14. Kingma, D.P.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
15. Laine, S., Hellsten, J., Karras, T., Seol, Y., Lehtinen, J., Aila, T.: Modular primitives for high-performance differentiable rendering. *ACM Transactions on Graphics* **39**(6) (2020)

16. Li, Z., Tucker, R., Snavely, N., Holynski, A.: Generative image dynamics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
17. Liang, H., Yin, Y., Xu, D., Liang, H., Wang, Z., Plataniotis, K.N., Zhao, Y., Wei, Y.: Diffusion4d: Fast spatial-temporal consistent 4d generation via video diffusion models. arXiv preprint arXiv:2405.16645 (2024)
18. Lin, Y., Dai, Z., Zhu, S., Yao, Y.: Gaussian-flow: 4d reconstruction with dynamic 3d gaussian particle. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21136–21145 (2024)
19. Liu, I., Su, H., Wang, X.: Dynamic gaussians mesh: Consistent mesh reconstruction from dynamic scenes. In: International Conference on Learning Representations. vol. 2025, pp. 1897–1932 (2025)
20. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: Seminal graphics: pioneering efforts that shaped the field, pp. 347–353 (1998)
21. Lu, C.Y., Zhou, P., Xing, A., Pokhariya, C., Dey, A., Shah, I.N., Mavidipalli, R., Hu, D., Comport, A.I., Chen, K., 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)
22. Mao, W., Hartley, R., Salzmann, M., et al.: Neural sdf flow for 3d reconstruction of dynamic scenes. In: The Twelfth International Conference on Learning Representations (2024)
23. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021)
24. Munkberg, J., Hasselgren, J., Shen, T., Gao, J., Chen, W., Evans, A., Müller, T., Fidler, S.: 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)
25. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5865–5874 (2021)
26. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: 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)
27. Rabich, S., Stotko, P., Klein, R.: Fpo++: efficient encoding and rendering of dynamic neural radiance fields by analyzing and enhancing fourier plenotrees. *The Visual Computer* **40**(7), 4777–4788 (2024)
28. Shao, R., Zheng, Z., Tu, H., Liu, B., Zhang, H., Liu, Y.: Tensor4d: Efficient neural 4d decomposition for high-fidelity dynamic reconstruction and rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16632–16642 (2023)
29. Shen, T., Gao, J., Yin, K., Liu, M.Y., Fidler, S.: Deep marching tetrahedra: a hybrid representation for high-resolution 3d shape synthesis. In: Advances in Neural Information Processing Systems (NeurIPS) (2021)
30. Shen, T., Munkberg, J., Hasselgren, J., Yin, K., Wang, Z., Chen, W., Gojcic, Z., Fidler, S., Sharp, N., Gao, J.: Flexible isosurface extraction for gradient-based mesh optimization. *ACM Trans. Graph.* **42**(4) (jul 2023). <https://doi.org/10.1145/3592430>

31. Shi, L., Hassanieh, H., Davis, A., Katabi, D., Durand, F.: Light field reconstruction using sparsity in the continuous fourier domain. *ACM Transactions on Graphics (TOG)* **34**(1), 1–13 (2014)
32. Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems* **33**, 7537–7547 (2020)
33. Wang, L., Zhang, J., Liu, X., Zhao, F., Zhang, Y., Zhang, Y., Wu, M., Yu, J., Xu, L.: 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)
34. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: 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*, Curran Associates Inc., Red Hook, NY, USA (2021)
35. Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L.: Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 3295–3306 (2023)
36. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
37. Xu, J., Fan, Z., Yang, J., Xie, J.: Grid4d: 4d decomposed hash encoding for high-fidelity dynamic gaussian splatting. *Advances in Neural Information Processing Systems* **37**, 123787–123811 (2024)
38. Yang, Z., Gao, X., Zhou, W., Jiao, S., Zhang, Y., Jin, X.: 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)
39. Ye, K., Gao, C., Li, G., Chen, W., Chen, B.: Geosplatting: Towards geometry guided gaussian splatting for physically-based inverse rendering. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 28991–29000 (2025)
40. Yu, Z., Peng, S., Niemeyer, M., Sattler, T., Geiger, A.: Monosdf: Exploring monocular geometric cues for neural implicit surface reconstruction. *Advances in neural information processing systems* **35**, 25018–25032 (2022)
41. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: 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)
42. Zhang, S., Wu, G., Xie, Z., Wang, X., Feng, B., Liu, W.: Dynamic 2d gaussians: Geometrically accurate radiance fields for dynamic objects. In: *Proceedings of the 33rd ACM International Conference on Multimedia*. pp. 8144–8153 (2025)
43. Zheng, C., Xue, L., Zarate, J., Song, J.: Gaustar: Gaussian surface tracking and reconstruction. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 16543–16553 (2025)