Disentangled 3D Gaussian Splatting: High-Fidelity Relightable Volumetric Video through Geometry-Appearance Decoupling

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 1404

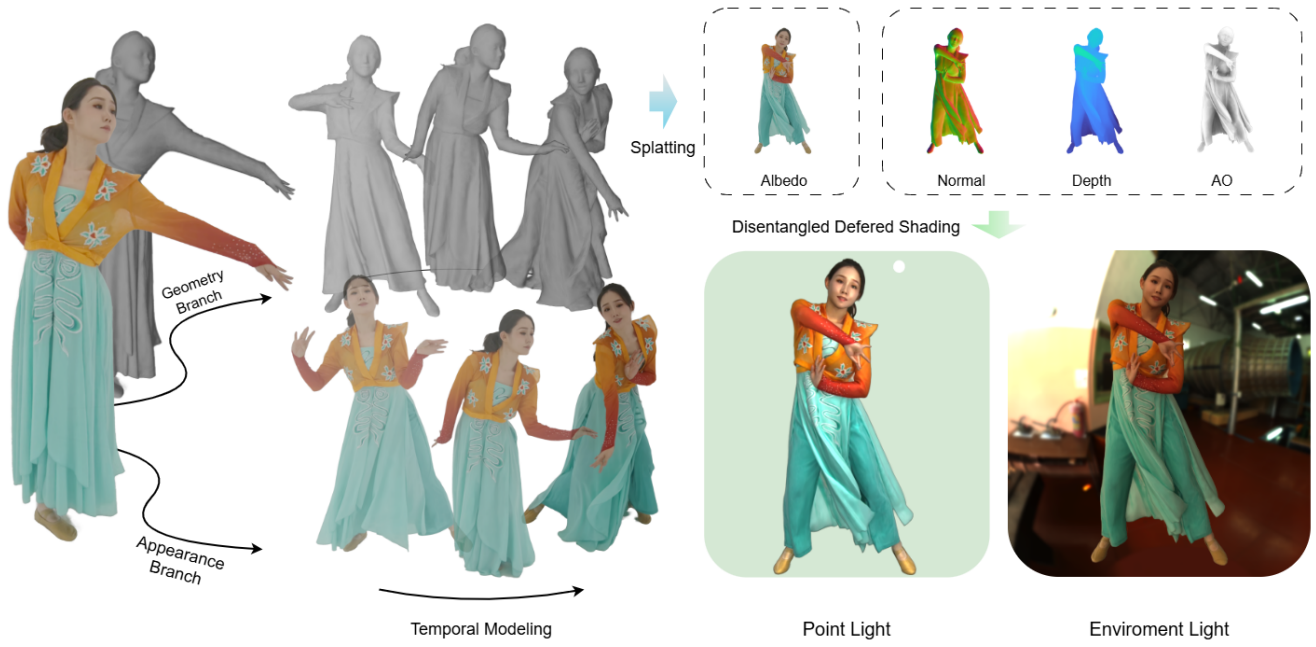

Fig. 1. We present a novel disentangled 3D Gaussian representation and extend this framework to temporal modeling, achieving both high-fidelity geometry reconstruction and photorealistic novel view synthesis. Our framework generates relightable volumetric video that supports practical downstream lighting manipulation via a deferred shading pipeline.

Recent advances in novel view synthesis have been greatly accelerated by the efficient, photorealistic rendering capabilities of 3D Gaussian Splatting (3DGS), bringing volumetric video technology closer to practical adoption. Numerous studies have successfully extended this approach to 4D reconstruction, showcasing its significant potential for volumetric video production. However, to fully realize the capabilities of 3DGS-based volumetric video, enhancing its geometric fidelity and relighting performance remains a critical challenge. In this work, we demonstrate that a controlled disentanglement of geometry and appearance modeling in 3DGS mutually enhances both aspects, leading to optimized convergence. Motivated by this observation, we propose a novel disentangled 3DGS representation that introduces a disentangling factor to regulate the interaction between geometry and appearance. Combined with well-designed pruning and densification strategies, our method improves performance in both novel view synthesis and surface reconstruction. We incorporate normal and visibility as additional Gaussian attributes, optimized in a staged manner, enabling high-quality relighting through a deferred shading pipeline. Furthermore, we extend our framework to the temporal domain, incorporating customized optimization strategies to produce temporally stable volumetric videos with relighting capabilities. Extensive experiments on both custom and public datasets demonstrate that our approach achieves competitive results in rendering quality and geometric reconstruction compared to existing methods.

CCS Concepts: • **Human-centered computing** → **Mixed / augmented reality**; • **Computing methodologies** → **Reconstruction**.

Additional Key Words and Phrases: 3D Gaussian Splatting, 3D Reconstruction, Relightable voluemtric video

ACM Reference Format:
Anonymous Author(s). 2025. Disentangled 3D Gaussian Splatting: High-Fidelity Relightable Volumetric Video through Geometry-Appearance Decoupling. *ACM Trans. Graph.* 1, 1 (August 2025), 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM 0730-0301/2025/8-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

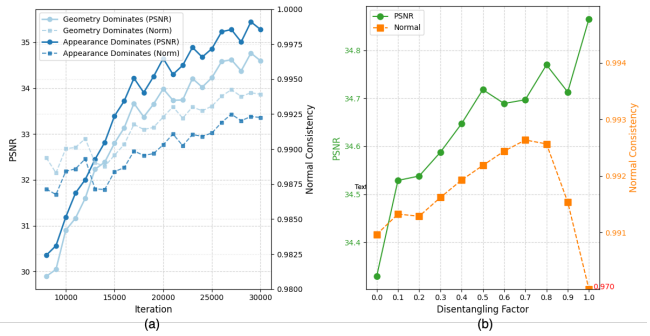

Fig. 2. Impact of the disentanglement in our approach. (a) demonstrates varying fitting abilities when emphasizing different branches through adjusted loss weights. (b) showing how the disentangling factor influences modeling quality.

1 INTRODUCTION

3D Gaussian Splatting (3DGS) [Kerbl et al. 2023] has emerged as a widely adopted representation for volumetric video production due to its photorealistic modeling capabilities. Unlike traditional mesh-based methods [Collet et al. 2015; Guo et al. 2019; Prada et al. 2017; Zhou et al. 2023], which often suffer from artifacts, low fidelity, and lack of realism, 3DGS demonstrates remarkable potential to capture fine geometric details. Its anisotropic modeling ability enables realistic rendering of reflective materials and mitigates multi-view parallax artifacts or blurring caused by inaccurate calibration while maintaining real-time rendering ability. Moreover, 3DGS excels at efficiently reconstructing intricate structures such as hair and fingers, which are critical yet challenging elements in human-centric volumetric applications. These advantages significantly expand its practical applicability and commercial value.

Recent advances have propelled volumetric video quality far beyond traditional mesh-based approaches. [Jiang et al. 2024b] pioneered sequential Gaussian modeling initialized from tracked meshes. [Wang et al. 2024] achieved high-compression-rate encoding for bandwidth-efficient streaming of 3DGS, [Xu et al. 2024] extended 4D reconstruction to long-duration sequences, and [Zhang et al. 2025] enabled unbounded-length volumetric video while preserving per-frame fidelity. Despite these innovations, key challenges remain for industrial adoption. A critical limitation is the lack of relightable modeling in existing methods, which severely restricts their utility for downstream applications. Incorporating relighting capabilities, including environment lighting and editable light sources, would substantially enhance immersion in virtual production pipelines.

Recent work has explored material parameterization in 3DGS for physics-based rendering (PBR). Methods like [Gao et al. 2024; Jiang et al. 2023; Shi et al. 2024; Wu et al. 2024a], have adopted deferred rendering strategies to decompose environment light and Gaussian PBR attributes through inverse rendering techniques. Through our experiments, we observed that naively incorporating geometry-related attributes into 3DGS is suboptimal, as it forces color and geometry to share the same distribution across the opacity field. We argue that these attributes should be disentangled: appearance exhibits higher-frequency features and greater tolerance to multi-view inconsistencies than geometry. Therefore, we conducted a series of experiments to maintain the mutual impact that boosts both geometry and appearance modeling, while enabling these two branches to achieve optimal distribution, resulting in better reconstruction results.

The most direct approach for disentangling geometry and appearance employs separate modeling frameworks. For instance, one might use a hybrid representation with meshes for geometry and 3DGS for rendering the appearance similar to [Cai et al. 2025; Lin et al. 2025]. Alternatively, [Zhou et al. 2025] using 2D Gaussian Splatting (2DGS) for geometry while modeling the appearance with 3DGS. However, for relighting applications, such hybrid representations often fail to maintain precise pixel-level alignment between geometry and appearance. Thus, a unified primitive that jointly models both aspects is essential.

We present a disentangled representation that independently parameterizes appearance and geometric attributes while preserving pixel-level alignment during rendering. Specifically, we disentangle appearance and geometry rendering into separate branches by disentangling geometric opacity fields from color opacity fields, accumulating them independently during splatting. We introduce a disentanglement factor to regulate the interaction between these branches, ensuring optimal convergence as shown in Fig. 2. Thanks to this representation, we can apply distinct regularization to each branch with minimal cross-influence.

By integrating this approach, we achieve stable and high-fidelity relightable 3DGS sequential modeling with temporal consistency. Our contributions include:

- A novel 3DGS-based representation that separates appearance and geometry attributes, mutually enhancing the quality of both branches.
- An improved training framework with refined pruning, densification strategies, and optimization schedules to produce compact surface representations with superior geometric and appearance fidelity for relightable 3DGS modeling.
- Our method extends to dynamic scenes, enabling high-fidelity, relightable 4D reconstruction with temporal coherence.

2 RELATED WORK

2.1 Volumetric Video

Volumetric video reconstruction traditionally builds upon novel view synthesis techniques. Early approaches [Collet et al. 2015; Prada et al. 2017; Zhou et al. 2023] employed mesh-based representations, where geometry was obtained through multi-view stereo reconstruction [Liu et al. 2010; Shao et al. 2022] or depth sensors [Lawrence et al. 2021; Newcombe et al. 2015; Yu et al. 2018; Zollhöfer et al. 2014]. These methods generated textured meshes by projecting multi-view imagery onto the reconstructed surfaces [Orts-Escolano et al. 2016]. However, they suffered from geometric inaccuracies, texture seams, ghosting artifacts, and topological inconsistencies during deformation resulting in rendering artifacts.

Neural radiance fields (NeRF) [Işık et al. 2023; Mildenhall et al. 2021; Müller et al. 2022; Park et al. 2021; Pumarola et al. 2020; Xian

et al. 2021] later addressed these issues, achieving photorealistic rendering quality. While NeRF-based methods significantly improved visual fidelity, they remained limited by implicit representations, leading to inefficient modeling and rendering. The advent of 3DGS [Kerbl et al. 2023; Lu et al. 2024] offered an explicit alternative that preserved rendering quality while improving practicality. Recent 4D extensions of 3DGS [Duan et al. 2024; Huang et al. 2023; Li et al. 2023; Luiten et al. 2024; Wu et al. 2024b; Yang et al. 2023] have demonstrated promising results for dynamic scenes. [Wang et al. 2024] proposes a group-of-pictures (GoP) modeling approach using MLPs with target compression along temporal Gaussians, and [Sun et al. 2024] introduces a two-stage training pipeline involving MLP-based deformation followed by per-frame refinement. [Jiang et al. 2024a; Xu et al. 2024; Zhang et al. 2025] extend these methods to support longer sequences while achieving higher reconstruction quality. Among these temporal modeling approaches, we build upon [Zhang et al. 2025] due to its per-frame refinement strategy, which effectively preserves high-frequency temporal details while retaining the representational benefits of 3DGS.

2.2 3DGS-based Geometry Modeling

Recent advances in 3DGS have focused mainly on enhancing geometric accuracy and surface reconstruction [Chen et al. 2024; Dai et al. 2024]. Notably, [Guédon and Lepetit 2024] proposed optimizing 3DGS to flatten primitives, encouraging the Gaussian field to approximate a zero-crossing of signed distance field (SDF). [Huang et al. 2024] introduced native 2D Gaussian Splatting (2DGS) for improved surface estimation.[Yu et al. 2024b], have developed methods for extracting well-defined zero-crossing surfaces from 3DGS, where surface normals are derived from the ray-Gaussian intersection plane. [Zhang et al. 2024a] further advanced geometric modeling through view-dependent distortion, although this introduced inconsistencies in normal image rasterization. Our experiments reveal that such geometric regularization often comes at the cost of appearance modeling flexibility. We found that extending normal attributes through splatting similar to [Gao et al. 2024] achieves comparable effectiveness without these limitations.

Subsequent work has significantly progressed in PBR modeling, with [Jiang et al. 2023; Keyang et al. 2024; Shi et al. 2024; Wu et al. 2024a] introducing efficient inverse rendering solutions, while [Moenne-Loccoz et al. 2024] incorporated ray tracing for further enhancement. These methods collectively progress toward joint geometry and material modeling, aligning with our objectives. However, these approaches typically depend on inverse rendering pipelines and precise environment map estimation [Munkberg et al. 2022; Verbin et al. 2022], which constrains their effectiveness for diffuse materials prevalent in volumetric video capture scenarios. Imperfect lighting estimation tends to amplify artifacts during temporal modeling. While high-fidelity material estimation typically requires programmable light stage [Guo et al. 2019; Saito et al. 2024], we adopt a simplified approach assuming constant environment lighting and diffuse materials with direct illumination modeling, enabling practical relighting through our geometric reconstruction framework.

3 PRELIMINARIES

3DGS represents a scene explicitly using anisotropic Gaussian ellipsoids. Each 3D Gaussian $G_i \in \{G_i | i = 1, \cdots, N\}$ is defined by a covariance matrix Σ_i and mean $\boldsymbol{\mu}_i$, along with three sets of spherical harmonics (SH) coefficients that describe its view-dependent color for each RGB channel. To be physically plausible, covariance matrix has to be positive semi-definite and therefore decomposed into a scaling matrix $\boldsymbol{S}$ and rotation matrix $\boldsymbol{R}$:

$$\Sigma_i = \boldsymbol{R}_i \boldsymbol{S}_i \boldsymbol{S}_i^\top \boldsymbol{R}_i^\top, \tag{1}$$

in which $\boldsymbol{S}_i$ is parameterized as a scaling vector $\boldsymbol{s}_i$ and $\boldsymbol{R}_i$ is parameterized as a quaternion $\boldsymbol{q}_i$.

When viewing from a specific camera perspective, a covariance matrix Σ is projected to the image space through:

$$\Sigma^{2D} = (\boldsymbol{J}\boldsymbol{W}\Sigma\boldsymbol{W}^\top\boldsymbol{J}^\top)_{1:2,1:2} \tag{2}$$

Here, $\boldsymbol{J}$ is the Jacobian of the affine approximation of the projective transformation and $\boldsymbol{W}$ denotes the view matrix of the camera. Meanwhile, the corresponding Gaussian mean $\boldsymbol{\mu}$ is projected onto the 2D plane resulting in $\boldsymbol{\mu}^{2D}$. With α_i^c in 2D plane computed from $\{\Sigma_i^{2D}, \boldsymbol{\mu}_i^{2D}\}$, the color $\mathcal{I}^c$ of a pixel can be retrieved by blending $\boldsymbol{\mathcal{N}}$ ordered Gaussian overlapping the pixel:

$$\mathcal{I}^c = \sum_{i=1}^{N} \boldsymbol{\psi}_i \alpha_i^c \prod_{j=1}^{i-1} (1 - \alpha_j^c) \tag{3}$$

where $\boldsymbol{\psi}_i$ represents the Gaussian attributes related to appearance modeling, comprising both the 0-degree spherical harmonic coefficient d_i^c and higher-order components d_i^r for view-dependent effects.

4 METHOD

Building upon dense multi-view RGB inputs, our pipeline first reconstructs the initial disentangled 3DGS representation, and then bakes the visibility to render ambiment occlusion (AO) maps, as illustrated in Fig. 3. We then extend this framework to sequential modeling based on [Zhang et al. 2025], enabling stable and explicit reconstruction of relightable 3DGS sequences.

4.1 Disentangled Modeling.

We enhance the standard 3DGS framework by introducing geometric parameters that undergo simultaneous rasterization with appearance attributes. Our key insight lies in the complete factorization of geometry and appearance through separate opacity fields: σ^g for geometry and σ^c for appearance. The geometric splatting operation is formulated as:

$$\mathcal{I}^g = \sum_{i=1}^{N} \boldsymbol{\gamma}_i \alpha_i^g \prod_{j=1}^{i-1} (1 - \alpha_j^g), \tag{4}$$

where $\boldsymbol{\gamma}_i$ represent geometric attributes including surface normal $\boldsymbol{n}_i$, and center position $\boldsymbol{\mu}_i$, visibility v_i, producing corresponding rasterized normal map $\boldsymbol{\mathcal{N}}$, and depth map $\mathcal{D}$ and visibility map $\mathcal{V}$. α_i^g is used for geometric blending.

Achieving complete disentanglement remains challenging, as geometric reconstruction fundamentally depends on the consistency of the multi-view appearance [Chen et al. 2024]. Our experimental

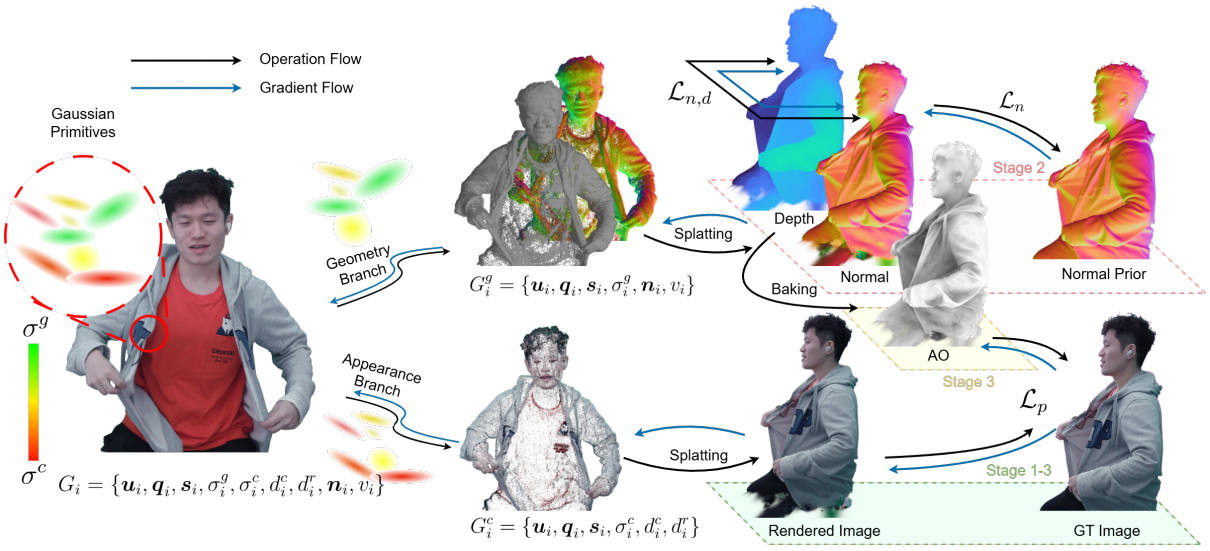

Fig. 3. Overview of our disentangled representation framework and training pipeline. The Gaussian primitives G_i are decomposed into geometry G_i^g and appearance G_i^c branches via a multi-stage optimization strategy. In Stage 1, we train a standard 3DGS model with zero-degree SH coefficients, focusing on optimizing the appearance branch to initialize the Gaussian distribution, with the disentanglement factor set to zero. In Stage 2, both the geometry and appearance branches are jointly optimized. Once convergence is achieved, AO is baked, and higher-order SH coefficients are further optimized in the final stage to achieve the complete model.

results reveal that joint optimization of both branches leads to a mutual improvement in performance, a finding consistent with [Zhou et al. 2025]. To precisely control this interdependence, we introduce disentanglement factor $\beta \in [0, 1]$ that balances the interaction between geometry and appearance:

$$\begin{aligned} \bar{\alpha}_i^c &= \frac{1}{2}\left((1+\beta)\alpha^c + (1-\beta)\alpha^g\right), \\ \bar{\alpha}_i^g &= \frac{1}{2}\left((1-\beta)\alpha^c + (1+\beta)\alpha^g\right). \end{aligned} \tag{5}$$

Here, $\bar{\alpha}_i^c$ and $\bar{\alpha}_i^g$ denote partially disentangled alpha values participating in the splatting process. As shown in Fig. 2(b), $\beta = 0$ merges the two branches, while $\beta = 1$ forces complete disentanglement.

Disentangled Densification. Similar to [Kerbl et al. 2023], we periodically densify underfitting regions to get better reconstruction. Our approach follows the standard 3DGS densification strategy, where both geometry and color branches determine densification based on pixel-space gradient magnitudes. For the opacity parameters geometry σ^g and appearance σ^c of each 3D Gaussian primitive, their influence on densification is quantified through:

$$\frac{\partial \mathcal{L}_i^c}{\partial \boldsymbol{u}_i^{2D}} \propto \frac{\partial \bar{\alpha}_i^c}{\partial \boldsymbol{u}_i^{2D}} \quad \text{and} \quad \frac{\partial \mathcal{L}_i^g}{\partial \boldsymbol{u}_i^{2D}} \propto \frac{\partial \bar{\alpha}_i^g}{\partial \boldsymbol{u}_i^{2D}}. \tag{6}$$

We monitor gradient statistics from both geometry and appearance branches to evaluate underfitting in local regions, triggering densification when these values exceed predefined thresholds. When a Gaussian is densified in one branch, while its counterpart in the other branch remains below the densification threshold, we initialize the new Gaussian in the inactive branch with minimal opacity. The opacity values for cloned Gaussians follow the revised formulation from [Bulò et al. 2024], ensuring stable training progress while maintaining the disentangling of geometric and appearance.

4.2 Relightable Geometric Surface Splatting

We reconstruct the geometry branch through a combination of inherent depth-normal consistency constraints and geometric priors, followed by ambient occlusion baking and higher-order spherical harmonics refinement. This pipeline ultimately enables practical relightable rendering through the deferred shading framework.

Geometric Splatting. Building on recent advances in vision foundation models [Khirodkar et al. 2024], we integrate monocular normal estimation as geometric guidance through a composite loss function:

$$\mathcal{L}_n = \lambda_n^{recon} L_1(\boldsymbol{N}, \hat{\boldsymbol{N}}) + \lambda_n^{smooth} \|\nabla \boldsymbol{N}\| \exp(-\|\nabla \hat{\boldsymbol{N}}\|). \tag{7}$$

The reconstruction term enforces alignment between our splatted normals $\boldsymbol{N}$ and the monocular prior $\hat{\boldsymbol{N}}$, while the edge-aware smoothing term preserves geometric discontinuities. Additionally, we employ a depth-normal consistency loss, similar to [Dai et al. 2024], to enforce consistency between normal and depth maps. This term also helps mitigate multi-view inconsistencies arising from monocular estimation:

$$\mathcal{L}_{n,d} = \lambda_n^{consist}(\text{sg}(O^g) - \boldsymbol{N} \cdot \frac{\nabla V(\mathcal{D})}{\|\nabla V(\mathcal{D})\|}). \tag{8}$$

Here, $V(\cdot)$ back-projects each pixel from the depth map to 3D space, producing a point map. O^g denotes the accumulated geometric alpha map, and $\text{sg}(\cdot)$ is the stop-gradient operator. During training, we do not normalize the output normal map, allowing it to softly

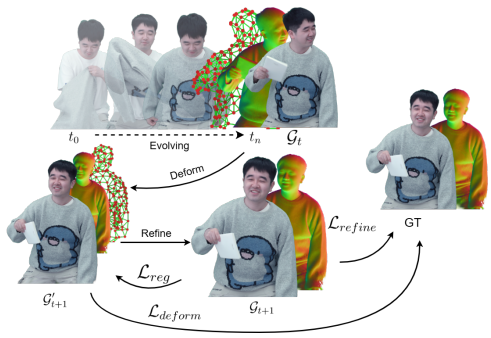

Fig. 4. Overview of our sequential modeling. The model evolves from the initial state t_0 to t_n through a two-stage reconstruction process, consisting of iterative deformation and refinement steps. During the deformation stage, the model is updated using the EDGraph, while the refinement stage closely follows the single-frame reconstruction pipeline.

align with the rendered anti-aliased color map. This approach yields improved relighting results at edges.

Baking Ambient Occlusion. The incident light from an environment for each direction ω_i is given by:

$$L_i(\omega_i) = V(\omega_i) \cdot L_{\text{direct}}(\omega_i) + L_{\text{indirect}}(\omega_i), \tag{9}$$

where $V(\omega_i)$ represents statistical visibility during ray tracing, L_{direct} denotes direct lighting, and L_{indirect} represents indirect lighting. Since volumetric video capture systems typically employ even lighting setups with low-reflection materials [Collet et al. 2015], we simplify this equation by ignoring indirect lighting. We model the view-dependent L_{direct} based on original high-order SH attributes d_i^r to capture isotropic effects under capture scenarios. For the ambient occlusion map, we first initialize visibility attributes for each 3DGS using point-based ray tracing with Bounding Volume Hierarchy (BVH), as suggested in [Gao et al. 2024]. To further bake the occlusion map and disentangle visibility shadows from color attributes, we utilize ray tracing for each shading point in training cameras. Based on the rasterized depth and normal maps, we generate downscaled and bilaterally filtered occlusion maps $\hat{\mathcal{V}}$. The rasterized AO map $\mathcal{V}$ serves as additional guidance to supervise visibility attributes:

$$\mathcal{L}_{\text{vis}} = \lambda_{\text{vis}}^{\text{recon}} L_1(\mathcal{V}, \hat{\mathcal{V}}) + \lambda_{\text{vis}}^{\text{smooth}} \|\nabla \mathcal{V}\| \exp(-\|\nabla \hat{\mathcal{N}}\|). \tag{10}$$

Since visibility edges are highly correlated with surface attributes, we use the prior normal map as guidance to achieve a smoother AO map.

Optimization. We employ a photometric loss $\mathcal{L}_p$ identical to [Kerbl et al. 2023], resulting in a full loss function:

$$\mathcal{L}_{recon} = \mathcal{L}_p + \mathcal{L}_n + \mathcal{L}_{n,d} + \mathcal{L}_{\text{vis}}. \tag{11}$$

We implement a three-stage computational pipeline to decompose surface attributes, as in Fig. 3. First, we fit 3DGS to obtain a rough point cloud reconstruction with base color attributes for initializing appearance and geometry modeling. Second, we training geometric branch for normal reconstruction and depth alignment. Finally, we bake visibility attributes and train for view-dependent effects.

4.3 Extending to Sequential Modeling

We extend our approach to sequential modeling following the method proposed in [Zhang et al. 2025], which produces a representation consistent with vanilla 3DGS. This approach implements a two-stage temporal reconstruction strategy: first deforming Gaussians set $\mathcal{G}_t$ to $\mathcal{G}'_{t+1}$ with an explicit embedding node graph (EDGraph) [Sumner et al. 2007] then performing refinement to get $\mathcal{G}_{t+1}$. This workflow is iterating through frames to reconstruct the entire sequence, as illustrated in Fig 4.

Skinning and Deformation. Each embedding node is sampled from the existing Gaussian primitives as u^e, with parameters including a local rotation quaternion $\boldsymbol{q}^e$ and translation $\boldsymbol{t}^e$. The warping of Gaussian primitive follows:

$$\tilde{\boldsymbol{\mu}}_i = \sum_{j \in N(G_i)} w_j(G_i)[\boldsymbol{q}_j^e \otimes (\boldsymbol{\mu}_i - \boldsymbol{u}_j^e) + \boldsymbol{u}_j^e + \boldsymbol{t}_j^e] \tag{12}$$

where $N(G_i)$ denotes the set of K^g nearest control nodes to Gaussian G_i. The blending weight $w_j(G_i)$ is computed as $w_j(G_i) = (1 - \|\boldsymbol{\mu}_i - \boldsymbol{c}_j\|/d_{\max})^2$, where $d_{\max}$ is the maximum distance to the $(K^g + 1)$-nearest node. The weights are normalized to sum to one. The quaternion multiplication operator is denoted by $\otimes$. Following the principle of locality, each point is influenced only by nearby control nodes, and node deformations are linearly blended to ensure smooth transitions.

Similarly, the normal attributes are deformed as:

$$\tilde{\boldsymbol{n}}_i = \sum_{j \in N(G_i)} w_j(G_i)[\boldsymbol{R}_j \cdot \boldsymbol{n}_i] \tag{13}$$

where $\boldsymbol{R}_j$ is the rotation matrix derived from quaternion $\boldsymbol{q}_j$. The overall optimization objective for this stage is:

$$\mathcal{L}_{deform} = \mathcal{L}_p + \mathcal{L}_n + \mathcal{L}_{quat} + \mathcal{L}_{ARAP} \tag{14}$$

where $\mathcal{L}_{quat}$ softly regularizes the quaternion norm to enforce unit length, and $\mathcal{L}_{ARAP}$ is an as-rigid-as-possible (ARAP) term to encourages local rigidity.

Refinement Stage. With a well-initialized Gaussian cloud after deformation, we jointly refine all attributes to obtain high-fidelity Gaussians. Unlike the previous training stage, we leverage the correspondences established during deformation to apply temporal smoothing, thereby achieving more stable geometric reconstruction through temporal regularization. For each geometric attribute:

$$\mathcal{L}_{\text{reg}} = L_1(\gamma_i, \tilde{\gamma}_i), \tag{15}$$

where $\tilde{\gamma}_i$ denotes the geometric attributes after deformation stage. Since physically deforming Gaussian points can introduce significant distortions, we incorporate an anisotropic loss as proposed in [Feng et al. 2024]:

$$\mathcal{L}_{aniso} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \max\left\{ \frac{S_p^1}{S_p^2} - a, 0 \right\}. \tag{16}$$

S_p^1 and S_p^2 are the principal scaling factors, and a is a threshold parameter. The complete optimization objective for the refinement stage is:

$$\mathcal{L}_{refine} = \mathcal{L}_p + \mathcal{L}_n + \mathcal{L}_{n,d} + \mathcal{L}_{vis} + \mathcal{L}_{reg} + \mathcal{L}_{aniso} \tag{17}$$

Fig. 5. We compare our method with both sequential modeling and geometry-based approaches for qualitative analysis in Tab. 3, with extra Deformable3DGS [Yang et al. 2023]. Overall, our method demonstrates better rendering quality, robustness to topology changes, and more accurate geometry estimation compared to existing techniques.

Table 1. Quantitative comparison. We evaluate the novel view synthesis . For temporal reconstruction methods the sequence are segmented with GoP length 20 to ensure best results on all methods. As for 3DGS-like methods we reconstruct the estimated mesh by truncated SDF fusion proposed by [Yu et al. 2024b]

best second best

Method	Lecturer			Dancer			ActorsHQ					Modeling	
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	Normal ↑	C.D.×10^{-2}↓	Sequential	Geometry
HumanRF [Işık et al. 2023]	28.24	0.897	0.154	27.42	0.861	0.121	30.62	0.941	0.076	-	-	✓	×
ReRF [Wang et al. 2023b]	24.93	0.746	0.409	19.51	0.749	0.370	28.10	0.830	0.292	-	-	✓	×
SC-GS [Huang et al. 2023]	29.36	0.906	0.175	26.15	0.879	0.182	28.30	0.929	0.141	-	-	✓	×
STG [Li et al. 2023]	29.10	0.952	0.101	27.20	0.913	0.109	26.95	0.875	0.289	-	-	✓	×
4DGS [Wu et al. 2024b]	27.32	0.954	0.124	25.65	0.909	0.145	24.12	0.876	0.290	-	-	✓	×
3DGStream [Sun et al. 2024]	28.26	0.942	0.148	27.96	0.901	0.102	26.59	0.874	0.150	-	-	✓	×
V^3 [Wang et al. 2024]	30.40	0.960	0.082	29.02	0.925	0.075	28.20	0.925	0.160	-	-	✓	×
EvolvingGS [Zhang et al. 2025]	33.33	0.969	0.074	29.09	0.928	0.074	30.87	0.952	0.095	-	0.596	✓	×
Ours	33.90	0.975	0.066	30.09	0.938	0.068	33.09	0.963	0.072	0.973	0.182	✓	✓
Neus2 [Wang et al. 2023a]	31.55	0.957	0.069	27.88	0.881	0.092	30.89	0.922	0.106	0.985	0.440	✓	✓
GSurfels [Dai et al. 2024]	31.34	0.963	0.098	27.70	0.906	0.132	29.40	0.937	0.161	0.973	0.290	×	✓
2DGS [Huang et al. 2024]	30.76	0.964	0.080	26.99	0.916	0.096	24.72	0.870	0.190	0.925	1.110	×	✓
R3DG [Gao et al. 2024]	29.97	0.963	0.097	20.63	0.880	0.206	30.19	0.956	0.063	0.943	0.287	×	✓
GOF [Yu et al. 2024b]	32.77	0.974	0.066	28.67	0.929	0.074	28.78	0.942	0.098	0.948	0.340	×	✓
RaDeGS [Zhang et al. 2024a]	33.01	0.975	0.069	28.84	0.931	0.070	28.99	0.948	0.093	0.957	0.216	×	✓
D2DGS [Zhang et al. 2024b]	29.84	0.831	0.125	28.07	0.862	0.158	27.73	0.920	0.161	0.956	0.350	✓	✓
Ours	33.90	0.975	0.066	30.09	0.938	0.068	33.09	0.963	0.072	0.973	0.182	✓	✓

5 IMPLEMENTATION

For basic reconstruction, we follow the hyperparameter settings from [Kerbl et al. 2023] and [Zhang et al. 2025]. We begin geometry training at 7,000 iterations and bake AO at 21,000 iterations. The weights for the objective function are set as follows: $\lambda_n^{recon} = 0.2$, $\lambda_n^{smooth} = 0.05$, $\lambda_n^{consist} = 0.01$, and $\lambda^{aniso} = 0.01$. The learning rate for normals is set to 0.01 by default. For the disentanglement strategy, we adopt a simulated annealing schedule, gradually increasing the disentanglement factor from 0 to 0.7 by the end of training. During the 3DGS Gaussian deformation stage, the regularization weights for each parameter are set as follows: 0.01 for geometry opacity, 1.0 for position, 0.3 for rotation, and 0.5 for normal attributes. The pruning threshold for ratio ρ_{max} is set to 0.25 for the ActorsHQ dataset and 0.1 for other datasets. We also employ a descending strategy to control the impact of pruning: specifically, we downscale this threshold by 0.75 each time the opacity is reset.

Disentangled Rasterization. Based on the vanilla 3DGS rasterizer, we implement disentangled branches by separating the alpha channel. To minimize the impact on rasterization efficiency and training time, geometry and color attributes are accumulated along the same rasterization pass with different weights during both the forward and backward passes. Additionally, we employ [Ye et al. 2024] to correct densification gradients and use the 2D box filter proposed in [Yu et al. 2024a] to alleviate zoom-out artifacts.

6 EXPERIMENTS

We evaluate our method on both public benchmarks and custom-captured datasets. The **ActorsHQ** dataset [Işık et al. 2023] consists of recordings of eight actors captured by 160 synchronized 12-megapixel cameras at 25 FPS, with ground-truth 4K meshes generated using multi-view geometry techniques. Ground-truth normal maps are rendered from the aligned RGB-mesh pairs. For evaluation, we downsample the input images to one-quarter of their original resolution to better assess mesh quality. In addition, we introduce two datasets from [Zhang et al. 2025] specifically designed to evaluate temporal performance under challenging real-world conditions. The **Dancer** dataset is captured using 52 10-megapixel cameras arranged in a three-tier configuration, while the **Lecturer** dataset utilizes 24 8-megapixel cameras focused on the subject's upper frontal body. Our custom datasets present significantly more challenging scenarios, featuring complex clothing, extreme motion, and substantial topological changes.

Table 2. Ablation study on ActorsHQ dataset. We evaluate the impact of our design on novel view synthesis ability and geometry quality.

	PSNR↑	SSIM↑	Normal↑	C.D.↓
baseline [Zhang et al. 2025]	30.30	0.950	-	0.534
w/o normal prior	32.09	0.951	0.952	0.274
w/o pruning strategy	31.96	0.956	0.954	0.239
w/o temporal regularization	32.53	0.957	0.973	0.213
w/o anisotropic regularization	32.37	0.955	0.972	0.238
Full	32.48	0.956	0.972	0.236

Quantitative and Qualitative Analysis. We conduct comparative evaluations with both temporal methods and single-frame reconstruction approaches to assess the quality of novel view synthesis and surface reconstruction accuracy for downstream relighting applications. For each sequence, we evaluate multiple clips using a GoP size of 20 frames. Using [Zhang et al. 2025] as our baseline, we compare our method against recent open-source 3DGS-based temporal modeling approaches, as shown in Tab. 3. By extending EvolvingGS's high-frequency detail recovery capability with our disentangled representation, our method achieves state-of-the-art results among temporal methods. In the second phase of evaluation, we compare with geometry-focused methods, particularly various 3DGS variants. While our method and D2DGS are evaluated using a GoP size of 20, other methods are assessed in a per-frame reconstruction manner. Our optimization strategy not only surpasses per-frame reconstruction in novel view synthesis tasks, but also produces high-quality surface splatting results on the ActorsHQ dataset. Selected qualitative comparisons are presented in Fig. 5.

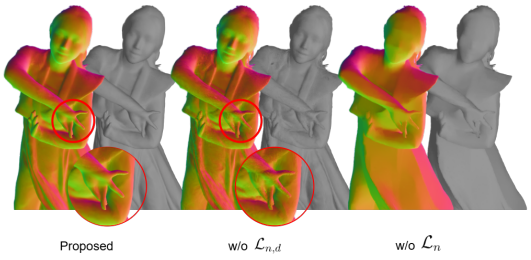

Fig. 6. Qualitative analysis of the ablation study on geometry branch reconstructions.

Fig. 7. Our reconstructed samples demonstrate robust interaction under various environmental lighting conditions.

Ablation Study. We evaluate our disentanglement framework on randomly sampled frames from our datasets. As shown in Fig. 2, appropriate tuning of the disentangling factor leads to significant improvements in both appearance and geometry modeling. Further comparisons, illustrated in Fig. 6, highlight the necessity of both our proposed normal prior injection for the reconstruction term $\mathcal{L}_n$ and the inherent depth-to-normal consistency term $\mathcal{L}_{n,d}$. The reconstruction prior provides detailed geometry, while the depth-to-normal consistency regularizes the opacity field and the positions of Gaussian points, preventing local noisy splatting artifacts. Furthermore, we present statistical results from our ablation studies in Tab. 2, which demonstrate the following: Normal prior supervision enables robust initialization and improves geometry reconstruction. The pruning strategy effectively removes local minima and redundant Gaussians, thereby enhancing the overall quality of reconstruction during temporal modeling, as also shown in Fig. **??**. Anisotropic regularization preserves the shape of primitives during deformation, leading to improved results. Temporal regularization trades a slight reduction in quality for increased stability across frames, which we recommend as the default setting. Overall, our configuration is compared to the baseline method EvolvingGS [Zhang et al. 2025], demonstrating the effectiveness of our disentangled framework.

7 DISCUSSIONS AND CONCLUSION

The core insight of our modeling approach is that splatting-based geometry representation requires separate accumulation processes for geometry and appearance attributes. By disentangling these two branches, we can apply distinct regularization strategies: for example, imposing stronger opacity constraints to preserve geometry, while avoiding unnecessary appearance restrictions that could blur high-frequency details or degrade novel view synthesis quality. This framework naturally accommodates branch-specific optimization strategies, including specialized pruning and densification for each domain.

Limitations. While our method builds upon [Zhang et al. 2025] and achieves superior reconstruction quality through the deform-then-refine paradigm, two limitations merit discussion. First, the approach demands greater computational resources due to its iterative per-frame optimization process. Second, unlike continuous 4D representations [Wu et al. 2024b], our method requires explicit regularization to maintain temporal coherence when handling complex topological changes and high-frequency details.

Future Work. The framework's modular design readily extends to physically-based rendering pipelines. As foundation models, like [Zeng et al. 2024], for material estimation advance, we anticipate seamless integration of PBR attribute decomposition via deferred rendering techniques. Such developments would not only enhance reconstruction fidelity but also bridge the gap to industrial applications in VR/AR content production, where material-accurate representations are crucial for photorealistic results under dynamic lighting conditions.

REFERENCES

Samuel Rota Bulò, Lorenzo Porzi, and Peter Kontschieder. 2024. Revising Densification in Gaussian Splatting. arXiv:2404.06109 [cs.CV] https://arxiv.org/abs/2404.06109

Hongrui Cai, Yuting Xiao, Xuan Wang, Jiafei Li, Yudong Guo, Yanbo Fan, Shenghua Gao, and Juyong Zhang. 2025. Hybrid Explicit Representation for Ultra-Realistic Head Avatars. arXiv:2403.11453 [cs.GR] https://arxiv.org/abs/2403.11453

Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. 2024. PGSR: Planar-based Gaussian Splatting for Efficient and High-Fidelity Surface Reconstruction. (2024).

Alvaro Collet, Ming Chuang, Pat Sweeney, Don Gillett, Dennis Evseev, David Calabrese, Hugues Hoppe, Adam Kirk, and Steve Sullivan. 2015. High-quality streamable free-viewpoint video. *ACM Transactions on Graphics* 34, 4 (2015), 1–13.

Pinxuan Dai, Jiamin Xu, Wenxiang Xie, Xinguo Liu, Huamin Wang, and Weiwei Xu. 2024. High-quality Surface Reconstruction using Gaussian Surfels. In *ACM SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery, Article 22, 11 pages.

Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 2024. 4D-Rotor Gaussian Splatting: Towards Efficient Novel View Synthesis for Dynamic Scenes. In *Proc. SIGGRAPH*.

Yutao Feng, Xiang Feng, Yintong Shang, Ying Jiang, Chang Yu, Zeshun Zong, Tianjia Shao, Hongzhi Wu, Kun Zhou, Chenfanfu Jiang, and Yin Yang. 2024. Gaussian Splashing: Unified Particles for Versatile Motion Synthesis and Rendering. *arXiv preprint arXiv:2401.15318* (2024).

Jian Gao, Chun Gu, Youtian Lin, Zhihao Li, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. 2024. Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *European Conference on Computer Vision*. Springer, 73–89.

Antoine Guédon and Vincent Lepetit. 2024. SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering. *CVPR* (2024).

Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, et al. 2019. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics* 38, 6 (2019), 1–19.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2024. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In *SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery. https://doi.org/10.1145/3641519.3657428

Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. 2023. SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes. *arXiv preprint arXiv:2312.14937* (2023).

Mustafa Işık, Martin Rünz, Markos Georgopoulos, Taras Khakhulin, Jonathan Starck, Lourdes Agapito, and Matthias Nießner. 2023. HumanRF: High-Fidelity Neural Radiance Fields for Humans in Motion. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–12. https://doi.org/10.1145/3592415

Yuheng Jiang, Zhehao Shen, Yu Hong, Chengcheng Guo, Yize Wu, Yingliang Zhang, Jingyi Yu, and Lan Xu. 2024a. Robust Dual Gaussian Splatting for Immersive Human-centric Volumetric Videos. *arXiv preprint arXiv:2409.08353* (2024).

Yuheng Jiang, Zhehao Shen, Penghao Wang, Zhuo Su, Yu Hong, Yingliang Zhang, Jingyi Yu, and Lan Xu. 2024b. HiFi4G: High-Fidelity Human Performance Rendering via

Compact Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 19734–19745.

Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. 2023. GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces. *arXiv preprint arXiv:2311.17977* (2023).

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023). https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/

Ye Keyang, Hou Qiming, and Zhou Kun. 2024. 3D Gaussian Splatting with Deferred Reflection. (2024).

Rawal Khirodkar, Timur Bagautdinov, Julieta Martinez, Su Zhaoen, Austin James, Peter Selednik, Stuart Anderson, and Shunsuke Saito. 2024. Sapiens: Foundation for Human Vision Models. *arXiv preprint arXiv:2408.12569* (2024).

Jason Lawrence, Dan B Goldman, Supreeth Achar, Gregory Major Blascovich, Joseph G. Desloge, Tommy Fortes, Eric M. Gomez, Sascha Häberling, Hugues Hoppe, Andy Huibers, Claude Knaus, Brian Kuschak, Ricardo Martin-Brualla, Harris Nover, Andrew Ian Russell, Steven M. Seitz, and Kevin Tong. 2021. Project Starline: A high-fidelity telepresence system. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)* 40(6) (2021).

Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. 2023. Spacetime Gaussian Feature Splatting for Real-Time Dynamic View Synthesis. *arXiv preprint arXiv:2312.16812* (2023).

Ancheng Lin, Yusheng Xiang, Paul Kennedy, and Jun Li. 2025. Direct Learning of Mesh and Appearance via 3D Gaussian Splatting. arXiv:2405.06945 [cs.CV] https://arxiv.org/abs/2405.06945

Yebin Liu, Qionghai Dai, and Wenli Xu. 2010. A Point-Cloud-Based Multiview Stereo Algorithm for Free-Viewpoint Video. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (May 2010), 407–418. https://doi.org/10.1109/TVCG.2009.88

Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. 2024. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20654–20664.

Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. 2024. Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis. In *3DV*.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.

Nicolas Moenne-Loccoz, Ashkan Mirzaei, Or Perel, Riccardo de Lutio, Janick Martinez Esturo, Gavriel State, Sanja Fidler, Nicholas Sharp, and Zan Gojcic. 2024. 3D Gaussian Ray Tracing: Fast Tracing of Particle Scenes. *ACM Transactions on Graphics and SIGGRAPH Asia* (2024).

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics* 41, 4 (2022), 1–15.

Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Müller, and Sanja Fidler. 2022. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8280–8290.

Richard A Newcombe, Dieter Fox, and Steven M Seitz. 2015. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 343–352.

Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-Time. In *UIST 2016*.

Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. 2021. HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields. *ACM Trans. Graph.* 40, 6, Article 238 (dec 2021).

Fabián Prada, Misha Kazhdan, Ming Chuang, Alvaro Collet, and Hugues Hoppe. 2017. Spatiotemporal atlas parameterization for evolving meshes. *ACM Trans. Graph.* 36, 4, Article 58 (July 2017), 12 pages. https://doi.org/10.1145/3072959.3073679

Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. 2020. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable Gaussian Codec Avatars. In *CVPR*.

Ruizhi Shao, Hongwen Zhang, He Zhang, Mingjia Chen, Yanpei Cao, Tao Yu, and Yebin Liu. 2022. DoubleField: Bridging the Neural Surface and Radiance Fields for High-fidelity Human Reconstruction and Rendering. In *CVPR*.

Yahao Shi, Yanmin Wu, Chenming Wu, Xing Liu, Chen Zhao, Haocheng Feng, Jian Zhang, Bin Zhou, Errui Ding, and Jingdong Wang. 2024. GIR: 3D Gaussian Inverse Rendering for Relightable Scene Factorization. arXiv:2312.05133 [cs.CV] https://arxiv.org/abs/2312.05133

Robert W. Sumner, Johannes Schmid, and Mark Pauly. 2007. Embedded deformation for shape manipulation. *ACM Trans. Graph.* 26, 3 (jul 2007), 80–es. https://doi.org/10.1145/1276377.1276478

Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, and Wei Xing. 2024. 3DGStream: On-the-Fly Training of 3D Gaussians for Efficient Streaming of Photo-Realistic Free-Viewpoint Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20675–20685.

Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. 2022. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. *CVPR* (2022).

Liao Wang, Qiang Hu, Qihan He, Ziyu Wang, Jingyi Yu, Tinne Tuytelaars, Lan Xu, and Minye Wu. 2023b. Neural Residual Radiance Fields for Streamably Free-Viewpoint Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 76–87.

Penghao Wang, Zhirui Zhang, Liao Wang, Kaixin Yao, Siyuan Xie, Jingyi Yu, Minye Wu, and Lan Xu. 2024. Viewing Volumetric Videos on Mobiles via Streamable 2D Dynamic Gaussians. arXiv:2409.13648 [cs.CV] https://arxiv.org/abs/2409.13648

Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. 2023a. NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 2024b. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20310–20320.

Tong Wu, Jia-Mu Sun, Yu-Kun Lai, Yuewen Ma, Leif Kobbelt, and Lin Gao. 2024a. DeferredGS: Decoupled and Editable Gaussian Splatting with Deferred Shading. arXiv:2404.09412 [cs.CV] https://arxiv.org/abs/2404.09412

Wenqi Xian, Jia-Bin Huang, Johannes Kopf, and Changil Kim. 2021. Space-time Neural Irradiance Fields for Free-Viewpoint Video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9421–9431.

Zhen Xu, Yinghao Xu, Zhiyuan Yu, Sida Peng, Jiaming Sun, Hujun Bao, and Xiaowei Zhou. 2024. Representing Long Volumetric Video with Temporal Gaussian Hierarchy. *ACM Transactions on Graphics* 43, 6 (November 2024). https://zju3dv.github.io/longvolcap

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. 2023. Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction. *arXiv preprint arXiv:2309.13101* (2023).

Zongxin Ye, Wenyu Li, Sidun Liu, Peng Qiao, and Yong Dou. 2024. AbsGS: Recovering Fine Details for 3D Gaussian Splatting. arXiv:2404.10484 [cs.CV]

Tao Yu, Zerong Zheng, Kaiwen Guo, Jianhui Zhao, Qionghai Dai, Hao Li, Gerard Pons-Moll, and Yebin Liu. 2018. DoubleFusion: Real-Time Capture of Human Performances with Inner Body Shapes from a Single Depth Sensor. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7287–7296. https://doi.org/10.1109/CVPR.2018.00761

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. 2024a. Mip-Splatting: Alias-free 3D Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 19447–19456.

Zehao Yu, Torsten Sattler, and Andreas Geiger. 2024b. Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes. *ACM Transactions on Graphics* (2024).

Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. 2024. RGBX: Image decomposition and synthesis using material- and lighting-aware diffusion models. In *ACM SIGGRAPH 2024 Conference Papers* (Denver, CO, USA) *(SIGGRAPH '24)*. Association for Computing Machinery, New York, NY, USA, Article 75, 11 pages. https://doi.org/10.1145/3641519.3657445

Baowen Zhang, Chuan Fang, Rakesh Shrestha, Yixun Liang, Xiaoxiao Long, and Ping Tan. 2024a. RaDe-GS: Rasterizing Depth in Gaussian Splatting. arXiv:2406.01467 [cs.GR] https://arxiv.org/abs/2406.01467

Chao Zhang, Yifeng Zhou, Shuheng Wang, Wenfa Li, Degang Wang, Yi Xu, and Shaohui Jiao. 2025. EvolvingGS: High-Fidelity Streamable Volumetric Video via Evolving 3D Gaussian Representation. arXiv:2503.05162 [cs.CV] https://arxiv.org/abs/2503.05162

Shuai Zhang, Guanjun Wu, Xinggang Wang, Bin Feng, and Wenyu Liu. 2024b. Dynamic 2D Gaussians: Geometrically accurate radiance fields for dynamic objects. arXiv:2409.14072 [cs.CV] https://arxiv.org/abs/2409.14072

Qingyuan Zhou, Yuehu Gong, Weidong Yang, Jiaze Li, Yeqi Luo, Baixin Xu, Shuhao Li, Ben Fei, and Ying He. 2025. MGSR: 2D/3D Mutual-boosted Gaussian Splatting for High-fidelity Surface Reconstruction under Various Light Conditions. arXiv:2503.05182 [cs.CV] https://arxiv.org/abs/2503.05182

Yifeng Zhou, Shuheng Wang, Wenfa Li, Chao Zhang, Li Rao, Pu Cheng, Yi Xu, Jinle Ke, Wenduo Feng, Wen Zhou, Hao Xu, Yukang Gao, Yang Ding, Weixuan Tang, and Shaohui Jiao. 2023. Live4D: A Real-time Capture System for Streamable Volumetric Video. In *SIGGRAPH Asia 2023 Technical Communications*. 1–4.

Michael Zollhöfer, Matthias Nießner, Shahram Izadi, Christoph Rehmann, Christopher Zach, Matthew Fisher, Chenglei Wu, Andrew Fitzgibbon, Charles Loop, Christian Theobalt, et al. 2014. Real-time non-rigid reconstruction using an RGB-D camera. *ACM Transactions on Graphics (ToG)* 33, 4 (2014), 1–12.

Table 3. Quantitative comparison. We evaluate the novel view synthesis . For temporal reconstruction methods the sequence are segmented with GoP length 20 to ensure best results on all methods. As for 3DGS-like methods we reconstruct the estimated mesh by truncated SDF fusion proposed by [Yu et al. 2024b]

best second best

Method	Lecturer			Dancer			ActorsHQ					Modeling	
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	Normal ↑	C.D.×10^{-2}↓	Sequential	Geometry
HumanRF [Işık et al. 2023]	28.24	0.897	0.154	27.42	0.861	0.121	30.62	0.941	0.076	-	-	✓	×
ReRF [Wang et al. 2023b]	24.93	0.746	0.409	19.51	0.749	0.370	28.10	0.830	0.292	-	-	✓	×
SC-GS [Huang et al. 2023]	29.36	0.906	0.175	26.15	0.879	0.182	28.30	0.929	0.141	-	-	✓	×
STG [Li et al. 2023]	29.10	0.952	0.101	27.20	0.913	0.109	26.95	0.875	0.289	-	-	✓	×
4DGS [Wu et al. 2024b]	27.32	0.954	0.124	25.65	0.909	0.145	24.12	0.876	0.290	-	-	✓	×
3DGStream [Sun et al. 2024]	28.26	0.942	0.148	27.96	0.901	0.102	26.59	0.874	0.150	-	-	✓	×
V^3 [Wang et al. 2024]	30.40	0.960	0.082	29.02	0.925	0.075	28.20	0.925	0.160	-	-	✓	×
EvolvingGS [Zhang et al. 2025]	33.33	0.969	0.074	29.09	0.928	0.074	30.87	0.952	0.095	-	0.596	✓	×
Ours	33.90	0.975	0.066	30.09	0.938	0.068	33.09	0.963	0.072	0.973	0.182	✓	✓
Neus2 [Wang et al. 2023a]	31.55	0.957	0.069	27.88	0.881	0.092	30.89	0.922	0.106	0.985	0.440	✓	✓
GSurfels [Dai et al. 2024]	31.34	0.963	0.098	27.70	0.906	0.132	29.40	0.937	0.161	0.973	0.290	×	✓
2DGS [Huang et al. 2024]	30.76	0.964	0.080	26.99	0.916	0.096	24.72	0.870	0.190	0.925	1.110	×	✓
R3DG [Gao et al. 2024]	29.97	0.963	0.097	20.63	0.880	0.206	30.19	0.956	0.063	0.943	0.287	×	✓
GOF [Yu et al. 2024b]	32.77	0.974	0.066	28.67	0.929	0.074	28.78	0.942	0.098	0.948	0.340	×	✓
RaDeGS [Zhang et al. 2024a]	33.01	0.975	0.069	28.84	0.931	0.070	28.99	0.948	0.093	0.957	0.216	×	✓
D2DGS [Zhang et al. 2024b]	29.84	0.831	0.125	28.07	0.862	0.158	27.73	0.920	0.161	0.956	0.350	✓	✓
Ours	33.90	0.975	0.066	30.09	0.938	0.068	33.09	0.963	0.072	0.973	0.182	✓	✓

We introduce a disentanglement factor to regulate the interaction between these branches, ensuring optimal convergence as shown in Fig. 2.

A IMPLEMENTATION

For basic reconstruction, we follow the hyperparameter settings from [Kerbl et al. 2023] and [Zhang et al. 2025]. We begin geometry training at 7,000 iterations and bake AO at 21,000 iterations. The weights for the objective function are set as follows: $\lambda_n^{recon} = 0.2$, $\lambda_n^{smooth} = 0.05$, $\lambda_n^{consist} = 0.01$, and $\lambda^{aniso} = 0.01$. The learning rate for normals is set to 0.01 by default. For the disentanglement strategy, we adopt a simulated annealing schedule, gradually increasing the disentanglement factor from 0 to 0.7 by the end of training. During the 3DGS Gaussian deformation stage, the regularization weights for each parameter are set as follows: 0.01 for geometry opacity, 1.0 for position, 0.3 for rotation, and 0.5 for normal attributes. The pruning threshold for ratio ρ_{max} is set to 0.25 for the ActorsHQ dataset and 0.1 for other datasets. We also employ a descending strategy to control the impact of pruning: specifically, we downscale this threshold by 0.75 each time the opacity is reset.

Disentangled Rasterization. Based on the vanilla 3DGS rasterizer, we implement disentangled branches by separating the alpha channel. To minimize the impact on rasterization efficiency and training time, geometry and color attributes are accumulated along the same rasterization pass with different weights during both the forward and backward passes. Additionally, we employ [Ye et al. 2024] to correct densification gradients and use the 2D box filter proposed in [Yu et al. 2024a] to alleviate zoom-out artifacts.