This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

ChronoGS: Disentangling Invariants and Changes in Multi-Period Scenes

Zhongtao Wang[1] Jiaqi Dai[2] Qingtian Zhu[3] Yilong Li[1] Mai Su[1]
Fei Zhu[1*] Meng Gai[1] Shaorong Wang[2] Chengwei Pan[4] Yisong Chen[1*] Guoping Wang[1]

[1] Peking University [2] Beijing Forestry University [3] The University of Tokyo [4] Beihang University

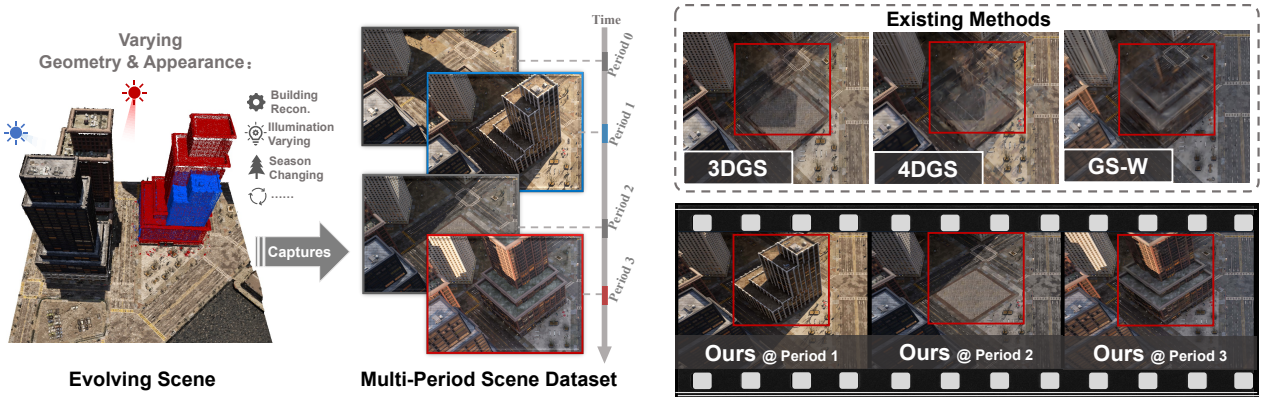


Figure 1. As time goes by, scenes naturally evolve — sunlight shifts, new structures emerge, seasons change, and so on. Cross-period captures therefore exhibit variations in both geometry and appearance. The red and blue overlays in the illustration simply exemplify the differences between the two periods. These coupled changes pose significant challenges for existing methods. Our approach, **ChronoGS**, robustly reconstructs and disentangles both geometry and appearance variations, delivering consistent and faithful results across time.

Abstract

*Multi-period image collections are common in real-world applications. Cities are re-scanned for mapping, construction sites are revisited for progress tracking, and natural regions are monitored for environmental change. Such data form multi-period scenes, where geometry and appearance evolve. Reconstructing such scenes is an important yet underexplored problem. Existing pipelines rely on incompatible assumptions: static and in-the-wild methods enforce a single geometry, while dynamic ones assume smooth motion, both failing under long-term, discontinuous changes. To solve this problem, we introduce **ChronoGS**, a temporally modulated Gaussian representation that reconstructs all periods within a unified anchor scaffold. It's also designed to disentangle stable and evolving components, achieving temporally consistent reconstruction of multi-period scenes. To catalyze relevant research, we release **ChronoScene** dataset, a benchmark of real and synthetic multi-period scenes, capturing geometric and appearance variation. Experiments demonstrate that ChronoGS consistently outperforms baselines in reconstruction quality and temporal consistency. Our code and the ChronoScene dataset are publicly available at https://github.com/ZhongtaoWang/ChronoGS.*

1. Introduction

Urban environments are periodically re-scanned for city mapping and management. Active construction sites are repeatedly surveyed to monitor progress and structural modifications. Regions affected by natural disasters are revisited over time to assess damage and document recovery. As a result, multi-period imagery of the same scene naturally accumulates—collected under different seasons, lighting, and even structural layouts. These multi-period captures contain valuable signals for long-term scene understanding, yet also bring unique challenges: geometry and appearance may evolve discontinuously due to construction, vegetation growth, or environmental variation.

Despite its clear importance, this multi-period scene reconstruction setting has remained largely unexplored. Existing pipelines are typically designed for two distinct set-

*Corresponding authors. Yisong Chen is the lead corresponding author.

tings. Static methods, including those designed for in-the-wild imagery, assume that all views share a time-invariant geometry. Dynamic methods, on the other hand, rely on smooth and continuous motion over time. Neither assumption holds when captures are seasons or even years apart. Consequently, current approaches fail to provide a consistent representation while flexibly modeling per-period variations. Our work aims to close this gap by introducing a framework that reconstructs all periods within a unified representation, disentangling what remains stable from what evolves.

We define a multi-period* scene as a set of captures of the same spatial region acquired at discrete time intervals, where the scene may have undergone geometric or appearance changes between captures. Our key observation is that multi-period scene reconstruction is neither static nor smoothly dynamic — it is discrete in periods but shared in major structure. The majority of spatial content remains invariant across periods, while changes are independent, enabling the scene to be decoupleable. This property allows the scene to be factorized into a shared canonical geometry and period-specific variations, which can be independently modeled within a unified framework.

Therefore, we propose **ChronoGS**, a temporally modulated Gaussian representation that models multi-period scenes consistently. ChronoGS reconstructs all multi-period scenes within a unified anchor scaffold of union geometry across all periods, providing a consistent geometric backbone for cross-period reconstruction. Within this scaffold, each anchor maintains an invariant base feature and a pool of local period-varying features that capture temporal variations. In addition, a global period-varying feature is adopted to model scene-level changes. These features allow anchors to vary over time while remaining a unified structure, ensuring stable and consistent reconstructions across periods. A geometry activation mechanism further enables the scaffold to express geometry specific to each period while maintaining global spatial consistency. Through this unified design, ChronoGS captures complex temporal phenomena, including both geometry and appearance change, achieving consistent reconstruction on multi-period scenes.

To support our research, we introduce **ChronoScene**, a novel benchmark designed to capture both geometry and appearance changes of multi-period scenes. Unlike existing benchmarks that assume static geometry or short-term continuous motion, this setting exposes the real challenges of large-scale scene evolution—building construction, vegetation growth, seasonal, and lighting shifts, etc. ChronoScene aligns multi-period captures in a common world coordinate frame, and provides a principled testbed for temporally modulated 3D modeling approaches.

Our experiments demonstrate that ChronoGS outperforms existing methods across diverse scenes, delivering better renders and more accurate geometry for each period while synthesizing plausible intermediate states. Ablation studies prove the rationality of our design. Also, we hope that ChronoGS and our dataset establish a foundation for future research on multi-period scene understanding.

Our contributions are threefold:

- **Novelty.** To the best of our knowledge, this is the **first** work to reconstruct multi-period scenes within a unified, differentiable Gaussian-splatting framework, jointly handling non-continuous geometry and appearance changes.
- **Dataset.** We introduce the **ChronoScene** dataset, a novel benchmark that aligns multi-period observations within a common coordinate and captures both geometric and appearance evolution across periods.
- **Method.** We propose **ChronoGS**, a temporally modulated Gaussian representation that effectively handles geometric and appearance variations that exist in multi-period scenes.

2. Related Work

2.1. Neural Radiance Fields

Neural rendering has advanced along two complementary tracks. Implicit approaches such as NeRFs [2, 22, 24, 25, 32] represent scenes as continuous volumetric fields, mapping 3D location and viewing direction to density and radiance via learned networks coupled with volumetric rendering. While achieving photorealistic novel views, dense ray sampling and repeated network evaluations are computationally expensive, prompting extensive work on acceleration [2, 6, 23]. In parallel, explicit point-based methods target real-time performance: 3D Gaussian Splatting [9] models scenes with rasterized Gaussian primitives for high-quality rendering, extended by hierarchical organization in Scaffold-GS [16], scalable variants for large scenes [10, 26, 31], and extract view-consistent surface representations [4, 8].

Despite these advances, both paradigms assume static scenes, forcing a single time-invariant representation of all observations. Jointly training all periods causes temporal averaging and artifacts, whereas per-period training avoids interference at the cost of heavier computation and storage without improving results. Our ChronoGS enables joint optimization across multiple periods, maintaining consistent long-term reconstruction in a unified model.

2.2. Dynamic Scene Representation

Building upon the success of NeRF in static scenes, several works have focused on extending it to dynamic en-

*We use the term **period** rather than **timestamp**. Because in our setting, each period corresponds to a time window during which images are captured at slightly different times. We manually define periods using metadata or prior knowledge, assuming that changes within a period are small compared to those across periods.

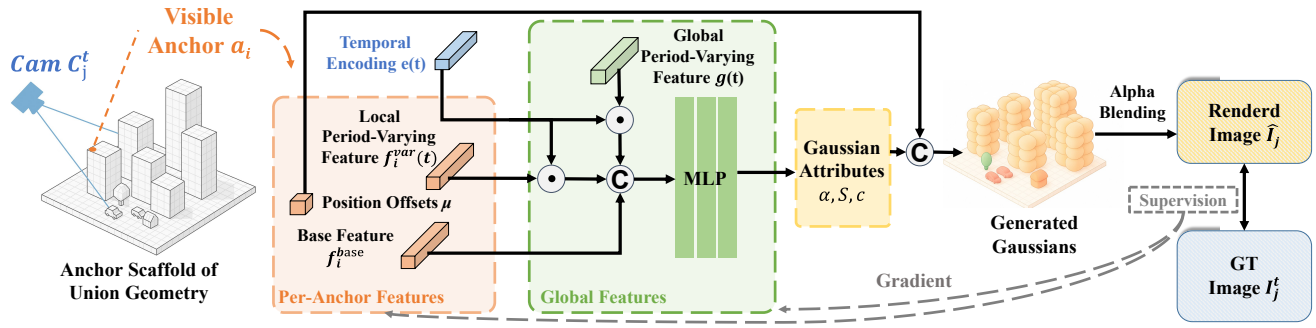


Figure 2. For a given camera C_j^t, we select visible anchors from our learned anchor scaffold of union geometry across periods. We use per-anchor's features and global features to render the multi-period scene. Temporal behavior is modeled via the anchor's local period-varying feature $f_i^{\text{var}}(t)$ and global period-varying feature $g(t)$ modulated by temporal encoding $e(t)$. After concatenating them with anchor's period-invariant base feature f_i^{base}, a lightweight MLP predicts Gaussian attributes $\{\alpha, S, c\}$. These attributes, along with the learned Gaussian position offset μ, produce a small cluster of Gaussians per anchor. We render the Gaussians generated by all visible anchors via differentiable splatting with alpha blending to obtain the image $\hat{I}_j^t$, and optimize with photometric losses between the ground truth image. This design enables consistent reconstruction of both geometry and appearance evolution on multi-period scenes.

vironments. D-NeRF [24] models motion by constructing a canonical space and applying deformation MLPs, although they are constrained by long-sequence videos. K-Planes [7] decompose features across multiple planes, enabling lightweight models to capture motion effectively. In parallel, 3DGS[9] has been extended to dynamic scenes. Dynamic 3DGS [17] optimizes Gaussians shared across time with regularizers, which rely on accurate point-cloud segmentation and additional temporal consistency terms. 4DGS [35] models dynamic scenes using canonical-space Gaussians combined with deformation MLPs. [5, 37] lift 3D Gaussians into 4D by adding a temporal dimension, and extract the corresponding 3D Gaussians at a given timestamp for rendering.

However, many of these methods assume temporally dense and smooth evolution, depend on precise masks or segmentations, and struggle with abrupt appearances and disappearances. These limitations are especially acute in multi-period data with long gaps and discontinuities. In contrast, our ChronoGS framework decouples the representation over periods, accommodating sudden shifts while preserving cross-period consistency.

2.3. Long-Term Scene Evolution Modeling

Several prior works study temporal evolution in visual data, but their goals differ from ours. Early approaches [21, 27, 29] infer temporal order from opportunistically captured Internet photos, while [28] emphasizes organization and visualization of historical photos with temporal metadata. Incremental reconstruction and time-lapse methods [18, 19, 33] build per-time updates or visualization outputs, rather than learning a unified renderable representation across periods.

Related *in-the-wild* novel view synthesis methods improve robustness to uncontrolled captures (unknown poses, illumination shifts, post-processing, and transient occluders). NeRF-W [20] models per-image appearance and transients, and NeRF-MS [12] regularizes appearance codes for cross-sequence consistency. NeuSC [14] addresses frequent appearance changes but assumes fixed geometry. In explicit representations, in-the-wild 3DGS methods [11, 15, 34, 36, 39] mainly target robust reconstruction under appearance variation and incremental updates. Continual adaptation methods such as [1, 3, 38] mitigate forgetting under incremental scene changes, but primarily capture relatively local and gradual geometry or appearance variation.

Overall, existing methods mostly focus on timeline inference, visualization, appearance variability, or incremental updating, and often collapse observations into a single state or assume fixed geometry. ChronoGS instead targets discontinuous multi-period captures with significant changes in both geometry and appearance, and learns a unified time-modulated representation that preserves cross-period consistency while enabling faithful period-specific rendering.

3. Method

Given a multi-period scene's captures $\{(\mathbf{I}_j^{(t)}, \mathbf{C}_j^{(t)})\}_{j=1}^{N}$ containing N images $\mathbf{I}_j^{(t)}$ and camera parameters $\mathbf{C}_j^{(t)}$ collected at $t = 1 \dots T$ discrete periods, our goal is to reconstruct all periods within a unified 3D representation. Different periods are often collected months or years apart, leading to substantial changes in both geometry and appearance.

We propose a temporally modulated Gaussian representation that can automatically distinguish period-varying and invariant components. Following the anchor-based paradigm of Scaffold-GS [16], we represent the scene with a set of *anchors* $\mathcal{A} = \{a_i\}_{i=1}^{N_a}$, each located at a fixed 3D

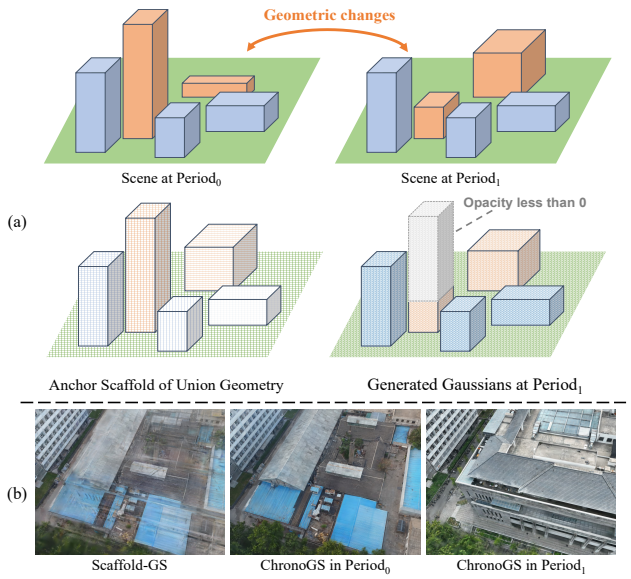


Figure 3. **Illustration of the temporal geometry activation mechanism. (a).** Anchor scaffold encodes the union of geometry across all periods, and generated gaussians with negative opacity are deactivated at certain periods, enabling the model to adaptively represent geometry variations over periods. **(b).** Comparison between Scaffold-GS and our ChronoGS trained on all periods' images. Scaffold-GS entangles structures from different periods, while ChronoGS successfully distinguishes and reconstructs period-specific geometry at Period$_0$ and Period$_1$.

position $\mathbf{x}_i$ and carrying learnable feature vectors. Given a camera, we select the visible anchors via frustum culling, and a lightweight MLP decodes each anchor's features into a small cluster of K 3D Gaussian primitives, which are used for differentiable splatting. The scaffold is initialized by voxelizing the merged sparse point clouds from all periods into a uniform spatial grid, so that both stable and varying geometry are covered from the start. Rather than training separate models per period, we optimize a single scaffold across periods, while temporal modulation and a geometry activation mechanism identify regions that change in appearance or geometry, enabling faithful per-period reconstruction and plausible intermediate-state synthesis. An overview is illustrated in Fig. 2.

3.1. Temporally Modulated Feature Representation

Anchor feature decomposition. To jointly represent shared and period-specific properties, each anchor a_i maintains two features: a period-invariant base feature $\mathbf{f}_i^{\text{base}} \in \mathbb{R}^{d_b}$ that describes shared geometry and appearance, and a local period-varying feature pool $\mathbf{f}_i^{var} = [\mathbf{f}_i^{(1)}, \ldots, \mathbf{f}_i^{(T)}] \in \mathbb{R}^{T \times d_v}$ that stores period-specific information.

Temporal Encoding. We represent temporal states using a unified T-dimensional encoding defined over period indices. For integer-valued period indices, we adopt a one-hot encoding where each observed period corresponds to a unique basis vector. For the inter-period position, we interpolate between the embeddings of adjacent periods to obtain a smooth representation:

$$\mathbf{e}(t) = (1-w)\mathbf{e}_{\lfloor t \rfloor} + w\mathbf{e}_{\lceil t \rceil}, \quad w = t - \lfloor t \rfloor, \qquad (1)$$

where $\mathbf{e}_{\lfloor t \rfloor}$ and $\mathbf{e}_{\lceil t \rceil}$ are the one-hot embeddings of neighboring periods. This formulation preserves exact encodings for observed periods while enabling smooth interpolation between adjacent periods.

Global–Local Temporal Modulation. The temporal encoding $\mathbf{e}(t)$ defined in Eq. (1) modulates both the local period-varying feature $\mathbf{f}_i^{\text{var}}(t)$, capturing per-anchor temporal variations, and a global period-varying feature $\mathbf{g}(t) \in \mathbb{R}^{T \times d_g}$ shared across all anchors, capturing scene-level factors such as illumination and seasonal appearance. Together with the period-invariant base feature $\mathbf{f}_i^{\text{base}}$, these three components allow the model to capture fine-grained dynamics while maintaining temporal consistency.

Explicit Gaussian Offsets for Geometric Stability. The i-th anchor stores K learnable but period-invariant Gaussian center offsets $\{\Delta\boldsymbol{\mu}_{ik}\}_{k=1}^{K}$, forming a fixed local arrangement around the anchor. Explicitly storing these offsets preserves local geometric stability across periods, so temporal geometry changes are mainly reflected through the Gaussians' visibility. This fixed spatial layout ensures consistent geometry across periods, enabling reliable opacity-based filtering described in Sec. 3.2.

3.2. Modeling Cross-Period Unified Geometry

Temporal geometry activation. As a unified framework, we design the anchor scaffold to align with the union geometry across all periods of the scene, and let the model learn which anchors should be activated at different periods. We model geometry as locally stable around each anchor, while temporal changes are captured by modulating their visibility. As illustrated in Fig. 3, to adaptively express the geometry of a specific period, we introduce a simple but effective geometry activation mechanism based on the predicted opacity. During rendering, Gaussian primitives whose decoded opacity $\alpha_{ik}(t)$ falls below zero are excluded from both compositing and gradient backpropagation. This mechanism automatically deactivates anchors corresponding to geometry that is occluded in certain periods. Consequently, the scaffold functions as a global geometric prior that dynamically activates relevant anchors per period, preserving both adaptability and structural consistency.

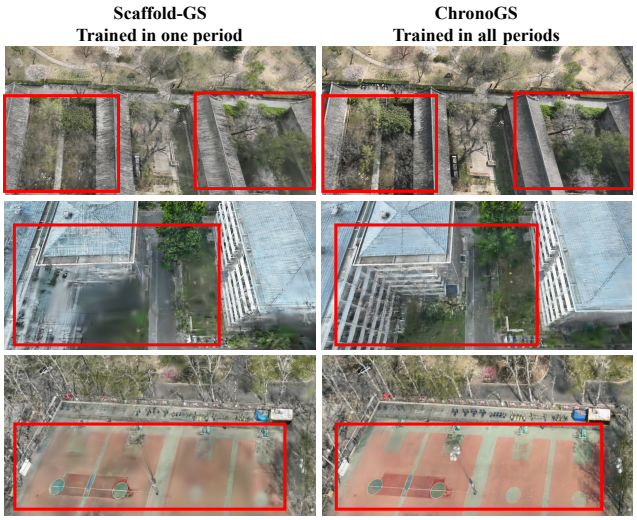


Figure 4. Comparison between Scaffold-GS trained in one period (left) and ChronoGS trained across all periods (right). Regions highlighted in red show that ChronoGS benefits from multi-period joint training, leveraging more observations to recover more accurate geometry and produce more complete reconstructions.

Cross-period benefit. Training on multi-period imagery in a unified anchor scaffold also benefits geometrically stable regions. It allows complementary supervision on all periods' capture to improve reconstruction quality. Anchors that persist across periods receive denser supervision from multi-period observations, leading to improved photometric consistency and reduced overfitting to single-period noise. This cross-period aggregation acts as an implicit regularizer: it encourages shared anchors to capture canonical geometry. As shown in Fig. 4, such joint optimization produces sharper textures and a more reliable geometry in areas that remain static across periods. Overall, the unified scaffold serves as a structural backbone that enables change geometry modeling. It also takes efficient use of all observations across periods in geometry-invariant areas, as opposed to training periods separately, which costs higher computation and storage.

3.3. Gaussian Decoding and Rendering

At each period t, we first fuse the features through temporal modulation:

$$\mathbf{h}_i(t) = \mathbf{g}(t) \odot \mathbf{e}(t) + \mathbf{f}_i^{\text{var}}(t) \odot \mathbf{e}(t) + \mathbf{f}_i^{\text{base}}, \quad (2)$$

where $\odot$ denotes channel-wise multiplication. Let $\mathbf{x}_i \in \mathbb{R}^3$ be the spatial position of anchor a_i, and $\mathbf{p}$ be the center of camera C_j^t. A lightweight MLP decoder Φ_θ maps $\mathbf{h}_i(t)$ to the parameters of K Gaussians:

$$\left\{\mathbf{S}_{ik}(t),\ \alpha_{ik}(t),\ \mathbf{c}_{ik}(t)\right\}_{k=1}^{K} = \Phi_\theta\big(\mathbf{h}_i(t),\ \widehat{\mathbf{d}}_i\big), \quad (3)$$

where $\mathbf{S}_{ik}(t)$ denotes the shape (rotation and scale), $\alpha_{ik}(t)$ the opacity, $\mathbf{c}_{ik}(t)$ the color of the k-th Gaussian associated with anchor a_i, and $\widehat{\mathbf{d}}_i = \mathbf{x}_i - \mathbf{p}$ is the viewing direction from the camera to the anchor. Each Gaussian's 3D center is obtained as $\boldsymbol{\mu}_{ik} = \mathbf{x}_i + \Delta\boldsymbol{\mu}_{ik}$, where $\Delta\boldsymbol{\mu}_{ik}$ is a learnable offset explicitly stored within anchor a_i.

Finally, the rendered color $\mathbf{C}(\mathbf{r}, t)$ for a camera ray $\mathbf{r}$ follows the standard volumetric alpha compositing of 3DGS [9]. We first project 3D Gaussians to 2D with mean $\widehat{\boldsymbol{\mu}}_{ik}$ and covariance $\widehat{\boldsymbol{\Sigma}}_{ik}$. Then, a front-to-back alpha compositing yields the rendered color:

$$\widehat{\mathbf{C}}(\mathbf{u}) = \sum_n \widehat{\mathbf{c}}_n(\mathbf{u})\widehat{\alpha}_n(\mathbf{u}) \prod_{j<n}(1 - \widehat{\alpha}_j(\mathbf{u})). \quad (4)$$

3.4. Optimization

We supervise the rendered images $\widehat{\mathbf{I}}$ with ground-truth observations $\mathbf{I}$ using a hybrid photometric objective:

$$\mathcal{L} = \lambda\|\widehat{\mathbf{I}} - \mathbf{I}\|_1 + (1-\lambda)(1 - \text{SSIM}(\widehat{\mathbf{I}}, \mathbf{I})). \quad (5)$$

Following [16], we track the accumulated gradient magnitude of each anchor feature over recent iterations. Anchors with persistently small gradient magnitude are pruned to remove redundancy and reduce memory footprint. We also monitor the view-space positional gradients of the Gaussians emitted from each anchor. When these gradients exceed a threshold, we spawn new anchors at the corresponding 3D locations obtained by the positions of Gaussians from old anchors. Both pruning and growth operations are triggered at regular training intervals, allowing the scaffold to refine its structural support to cover the union of geometry across all periods while maintaining a consistent canonical representation.

4. Experiments

4.1. Implementation Details

We set anchor's base feature dimension to $d_b = 16$, the local period-varying feature dimension to $d_v = 16$, and the global feature dimension to $d_g = 32$. For each anchor, the lightweight MLP decoder predicts $K = 10$ Gaussians. All experiments are trained for 40k iterations to ensure sufficient convergence on large-scale multi-period scenes. Details and ablations on hyperparameters and training strategy are provided in the supplementary material.

4.2. ChronoScene Dataset

We introduce *ChronoScene*, a multi-period dataset designed to challenge long-term 3D reconstruction under both *geometric* and *appearance* evolution. As illustrated in Fig. 5, scenes in ChronoScene exhibit coupled changes such as building construction or demolition, vegetation growth, and variations in season, lighting, and weather. ChronoScene contains a total of **8,891 images** across **12 scenes** and **42 sub-scenes** (each sub-scene corresponds to one specific

Real Scenes	Avg.	Lawncourt			Overstreet			BioBuilding			Canteen			Incubator			TriPlot		
Metrics	Mem.↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS[9]	1.01GB	18.29	0.4658	0.4862	18.79	0.6640	0.3627	17.37	0.5754	0.4402	19.64	0.5885	0.4150	15.97	0.6232	0.3874	20.17	0.6320	0.3793
Scaffold-GS[16]	0.36GB	16.27	0.3749	0.5363	15.20	0.5433	0.4390	16.09	0.4876	0.4923	17.70	0.5165	0.4528	15.72	0.5524	0.4291	18.04	0.5553	0.4280
GS-W[39]	0.23GB	20.33	0.4018	0.5638	21.64	0.6644	0.3974	21.15	0.5536	0.5053	21.24	0.5227	0.4780	22.39	0.6788	0.3922	22.57	0.6064	0.4436
NeuSC[14]	0.62GB	14.40	0.1394	0.6871	12.48	0.0523	0.6825	13.82	0.0367	0.6707	16.56	0.3740	0.6613	12.81	0.0254	0.6859	17.27	0.4450	0.6195
CLNeRF[3]	0.19GB	15.72	0.2148	0.7507	13.07	0.3851	0.7690	15.66	0.3501	0.7276	16.70	0.3168	0.7242	15.93	0.4333	0.7114	18.28	0.4089	0.6641
4DGS[35]	0.18GB	19.03	0.4674	0.6085	19.12	0.5815	0.5110	17.18	0.5013	0.6211	21.67	0.5932	0.5345	18.39	0.5538	0.5333	19.35	0.5093	0.5691
realtime4DGS[37]	5.53GB	20.91	0.5124	0.4943	17.17	0.5087	0.4993	21.39	0.5801	0.4920	22.49	0.6181	0.4292	22.71	0.7070	0.3801	22.86	0.6681	0.3982
Ours	0.52GB	22.16	0.6533	0.3390	22.66	0.7736	0.2838	22.87	0.7221	0.3274	25.15	0.7538	0.2647	23.79	0.7852	0.2721	24.16	0.7474	0.2901
Synthetic Scenes	Avg.	Aerial1			Aerial2			Aerial3			Aerial4			Street1			Street2		
Metrics	Mem.↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS[9]	1.05GB	20.79	0.7831	0.3179	19.95	0.7877	0.2252	23.20	0.8250	0.2682	21.41	0.7934	0.2631	15.40	0.6763	0.4065	18.77	0.7225	0.3580
Scaffold-GS[16]	0.30GB	19.52	0.7273	0.3605	18.35	0.7199	0.2738	19.51	0.7485	0.3327	18.38	0.7020	0.3272	14.49	0.5905	0.5252	16.86	0.5856	0.5279
GS-W [39]	0.22GB	25.11	0.7581	0.3754	23.21	0.7005	0.2510	28.79	0.7958	0.3186	26.13	0.7826	0.2644	15.73	0.6328	0.4847	18.93	0.7075	0.3873
NeuSC[14]	0.62GB	15.88	0.1929	0.6244	14.85	0.2385	0.6512	16.82	0.1996	0.5843	16.69	0.3160	0.6230	11.60	0.0778	0.6256	14.53	0.1955	0.6630
CLNeRF[3]	0.20GB	20.47	0.5892	0.6513	22.21	0.5254	0.5950	26.07	0.6850	0.5148	23.89	0.6045	0.5699	21.25	0.5866	0.6046	23.17	0.6169	0.5833
4DGS[35]	0.18GB	20.69	0.7140	0.4375	19.62	0.6325	0.4324	22.47	0.7359	0.3971	20.67	0.6675	0.4369	13.53	0.5440	0.5635	16.03	0.5310	0.5928
realtime4DGS[37]	7.41GB	22.29	0.7774	0.3351	20.04	0.7088	0.3268	23.93	0.8044	0.3075	21.60	0.7443	0.3341	18.88	0.7121	0.3944	20.53	0.7451	0.3514
Ours	0.65GB	28.80	0.8562	0.2509	26.50	0.8467	0.1854	30.11	0.8751	0.2266	28.10	0.8350	0.2305	24.70	0.8071	0.3241	26.56	0.8220	0.2946

Table 1. Quantitative comparison on ChronoScene dataset. The best, second-best, and third-best results in each column are highlighted by red, orange, and yellow, respectively. The upper part shows results on real scenes, and the lower part on synthetic scenes.

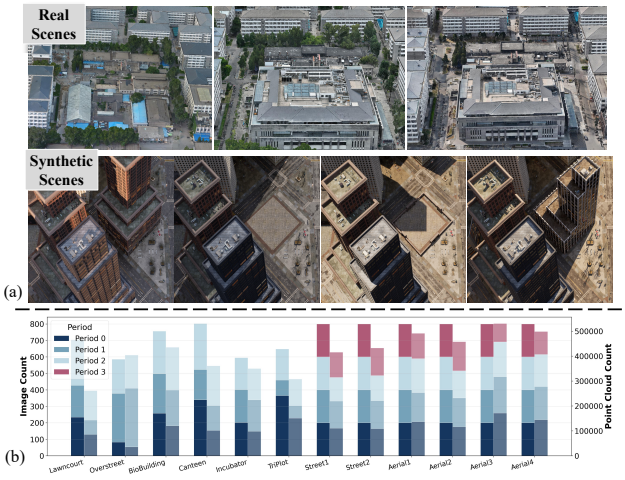


Figure 5. **(a). Examples from the ChronoScene dataset.** Top: real-world aerial captures showing natural evolution across long temporal intervals (e.g., construction and season change). Bottom: our synthetic scenes edited to simulate realistic changes with controllable geometry and appearance changes. **(b). Number distribution of image and point cloud across periods for 6 real-world and 6 synthetic scenes.** Left bars show image counts, right bars show sparse point cloud counts.

scene at one period). All images are registered into a *common world coordinate frame* in COLMAP [30] format, ensuring that cameras across periods are spatially comparable.

Real data. The real-world scenes span time gaps ranging from one year to three years between periods, reflecting natural changes in urban environments. Images are geo-tagged at capture time and globally aligned using GPS priors, followed by human-in-the-loop refinement to ensure consistent cross-period camera registration.

Synthetic data. To complement real data with controlled, precisely known temporal changes, we build synthetic scenes based on Matrix City [13]. We manually edit urban assets to simulate realistic evolutions—such as new constructions, façade refurbishments, or vegetation changes—and render both aerial and street-level views.

We release all RGB images, calibrated camera poses, and per-image period identifiers. More statistics, licensing, and details are provided in the supplementary material.

4.3. Evaluation

Unless otherwise noted, experiments in the main paper are conducted on **ChronoScene**, our primary benchmark for multi-period geometric and appearance evolution. To further assess generalization, we additionally report quantitative results on WAT[3], CL-Splat[1], and NeuSC[14] datasets in the supplementary material.

Baselines setup. We compare representative paradigms covering diverse settings: 3DGS [9] as the canonical static baseline, Scaffold-GS [16] introducing anchor-based structured Gaussians, GS-W [39] which targets uncontrolled photo collections, NeuSC [14] for long-term in-the-wild appearance changes, CLNeRF [3] for incremental scene updates, as well as 4DGS [35] and Realtime4DGS [37] that extend Gaussians to temporally continuous motion.

Protocol. Camera trajectories are independently captured at different periods, so we assess reconstruction quality *within each period*: for each test set of a given period, we render from its held-out cameras and compute PSNR, SSIM, and LPIPS with ground-truth images; metrics are then averaged over periods. We ensure that all methods use the same train-test splits and data augmentations.

4.4. Results

Quantitative performance. We report results separately for the real and synthetic portions of ChronoScene. Tab. 1 summarizes image-level metrics and overall storage cost in all 12 scenes of real and synthetic data, where geometry and appearance changes across periods. Our method consistently outperforms existing baselines on ChronoScene, achieving the best metrics across scenes while maintaining a compact memory footprint. These results confirm the ef-

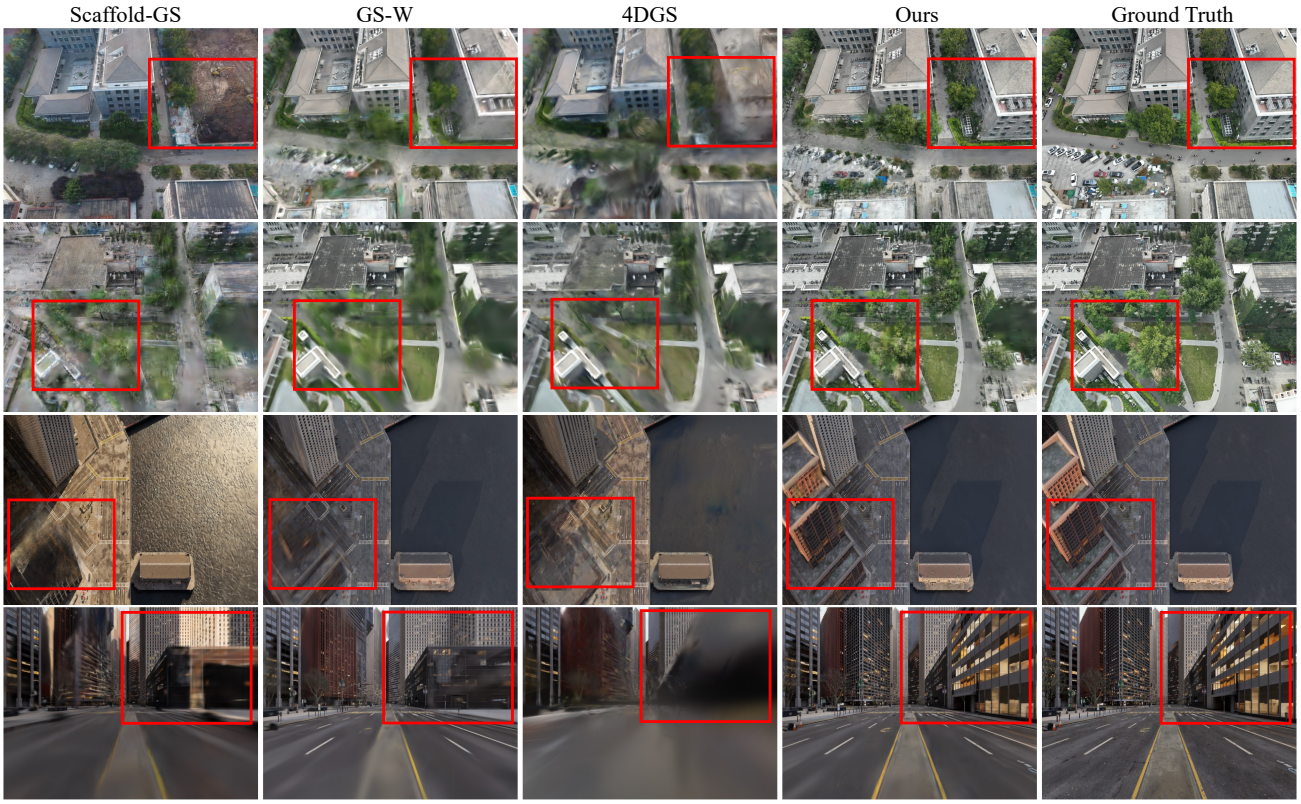


Figure 6. **Qualitative comparison on ChronoScene.** We compare ChronoGS with representative baselines. Each row shows novel-view renderings of different methods, along with ground truth. Static models trained on mixed multi-period data produce ghosting and appearance blending due to inconsistent geometry, while dynamic methods that assume continuous motion fail under the large temporal gaps, causing incorrect geometry interpolation. ChronoGS faithfully reconstructs period-specific geometry and appearance.

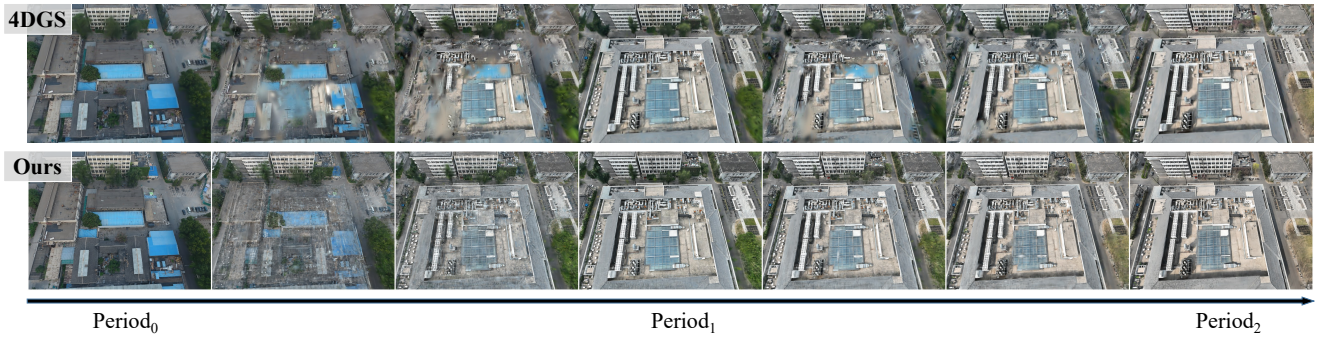


Figure 7. **Temporal evolution at a test viewpoint.** We query ChronoGS at continuous period indices $t \in [1, T]$ and render from the same camera. Compared to dynamic baselines such as 4DGS that assume smooth motion, which often hallucinate intermediate structures or blend neighboring periods, ChronoGS cleanly switches between periods, recovering accurate geometry and appearance at each period.

fectiveness of our method in modeling both geometry and appearance variations across periods.

Qualitative comparisons. Fig. 6 presents a side-by-side comparison of our renderings, ground truth, and representative baselines for held-out views in each period. Static models tend to blur conflicting configurations and dynamic models produce temporal smearing; in contrast, our approach maintains sharp, period-correct details. Our method preserves period-specific appearance and suppresses cross-period ghosting that appears in static models.

Temporal interpolation. To evaluate whether our model can synthesize intermediate periods, in Fig. 7 we show the renderings at the *same* test viewpoint across periods and compare our method to dynamic baselines. Methods designed for continuous dynamics often interpolate through non-existent intermediate states under ChronoScene's discontinuous region, leading to hallucinated transitions or residuals from neighboring periods. Our model cleanly

	3DGS [9]			Scaffold-GS [16]			Ours		
Period	PSNR↑	Mem.↓	Iters.	PSNR↑	Mem.↓	Iters.	PSNR↑	Mem.↓	Iters.
Period 0	21.97	1.6GB	40k	22.04	0.19GB	40k	**22.64**	N/A	N/A
Period 1	22.31	2.0GB	40k	21.88	0.25GB	40k	**22.50**	N/A	N/A
Period 2	22.23	1.3GB	40k	21.81	0.21GB	40k	**22.88**	N/A	N/A
Overall	22.23	4.9GB	120k	21.87	0.65GB	120k	**22.66**	**0.57GB**	**40k**

Table 2. Quantitative comparison of static methods' per-period independent training and our joint multi-period optimization on the *Overstreet* scene. Reported metrics include PSNR, memory footprint, and training iterations. ChronoGS achieves the **highest** quality across all periods while markedly reducing training iterations and storage requirements compared with static baselines.

	Lawncourt			Canteen		
Settings	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
w/o Var.&Global.	18.18	0.4609	0.4928	19.88	0.5959	0.4107
w/o Base.	21.60	0.6055	0.3918	24.57	0.7289	0.2970
w/o Var.	21.96	0.6440	0.3526	24.62	0.7426	0.2801
w/o Global.	22.11	0.6466	0.3497	24.94	0.7502	0.2716
Ours (full)	**22.16**	**0.6533**	**0.3390**	**25.15**	**0.7538**	**0.2647**

Table 3. Quantitative ablation results on two real scenes *Lawncourt* and *Canteen*. Removing the base, local period-varying, or global period-varying feature leads to performance degradation, confirming that all three are essential in our design.

switches between periods with more reasonable intermediate states.

Compare with Per-Period Training. To evaluate the effect of multi-period joint optimization, we compare ChronoGS with static baselines trained independently for each period under the same evaluation setting on *Overstreet* scene. As summarized in Tab. 2 and shown in Fig. 4, compared with training separately in each period's capture, our approach, which trains data from all periods together, receives denser and more complementary supervision—leading to the same or improved quality and significant reductions in both training and storage cost.

4.5. Ablation Study

To evaluate the role of each component in ChronoGS, we perform ablations under identical training settings. We compare the full model with several degraded variants: **w/o Base.**, removing the base feature $\mathbf{f}_i^{\text{base}}$ and forcing each anchor to rely only on period-dependent representations; **w/o Var.**, removing the local period-varying feature $\mathbf{f}_i^{\text{var}}(t)$ and disabling local temporal adaptation; **w/o Global.**, removing the global period-varying feature $\mathbf{g}(t)$ thus losing global appearance coherence; and **w/o Var.&Global.**, removing $\mathbf{g}(t)$ and $\mathbf{f}_i^{\text{var}}(t)$ to disable model's temporal behavior.

As shown in Tab. 3, the full ChronoGS achieves the best quantitative and perceptual scores. Our design disentangles changes from period-invariant structure to achieve consistent multi-period scene reconstruction. Without the base feature, the shared geometry becomes unstable. Without the local period-varying feature, local temporal changes cannot be expressed, and without the global period-varying feature, scene-level illumination and color consistency degrade noticeably. These can be clearly observed in Fig. 8.

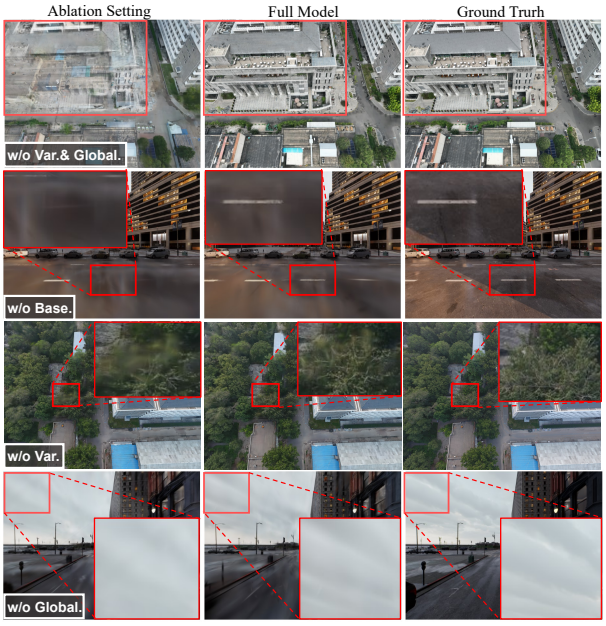


Figure 8. **Qualitative Ablation results.** Removing the base, local period-varying, or global period-varying feature leads to degraded geometry stability, local temporal fidelity, and scene-level appearance consistency, respectively, compared with the full model.

5. Conclusion

We presented **ChronoGS**, a temporally modulated Gaussian representation for reconstructing multi-period scenes. ChronoGS represents a multi-period scene with a unified anchor scaffold that aggregates geometry shared across periods. Temporally modulated features and a geometry activation mechanism jointly model both invariant structures and period-varying attributes. This formulation effectively handles geometric and appearance variations that arise when imagery is collected in different periods. In parallel, we introduced **ChronoScene**, a novel benchmark that collects multi-period observations and captures both geometric and appearance evolution across periods. Extensive experiments demonstrate that ChronoGS achieves sharper renderings, more accurate geometry, and improved cross-period consistency compared to existing baselines. Ablation studies validate the contribution of each component. To our knowledge, this is the first work to reconstruct multi-period scenes within a unified, differentiable Gaussian-splatting framework that jointly handles non-continuous geometry and appearance changes. We provide both a simple but strong baseline and a valuable public benchmark for future research.

Acknowledgments

This work is supported by the National Key R&D Program of China (Grant No. 2023YFB3309001). This work is also supported by the Shenzhen Science and Technology Program (Grant No. KJZD20230923114114028) and the Shandong Key Research and Development Program (Grant No. 2024CXGC010801).

References

[1] Jan Ackermann, Jonas Kulhanek, Shengqu Cai, Haofei Xu, Marc Pollefeys, Gordon Wetzstein, Leonidas J Guibas, and Songyou Peng. Cl-splats: Continual learning of gaussian splatting with local optimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7808–7817, 2025. 3, 6

[2] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479, 2022. 2

[3] Zhipeng Cai and Matthias Müller. Clnerf: Continual learning meets nerf. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23185–23194, 2023. 3, 6

[4] Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 2

[5] Yuanxing Duan, Fangyin Wei, Qiyu Dai, Yuhang He, Wenzheng Chen, and Baoquan Chen. 4d-rotor gaussian splatting: towards efficient novel view synthesis for dynamic scenes. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 3

[6] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5501–5510, 2022. 2

[7] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023. 3

[8] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH 2024 conference papers*, pages 1–11, 2024. 2

[9] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 3, 5, 6, 8

[10] Bernhard Kerbl, Andreas Meuleman, Georgios Kopanas, Michael Wimmer, Alexandre Lanvin, and George Drettakis. A hierarchical 3d gaussian representation for real-time rendering of very large datasets. *ACM Transactions on Graphics (TOG)*, 43(4):1–15, 2024. 2

[11] Jonas Kulhanek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. Wildgaussians: 3d gaussian splatting in the wild. *arXiv preprint arXiv:2407.08447*, 2024. 3

[12] Peihao Li, Shaohui Wang, Chen Yang, Bingbing Liu, Weichao Qiu, and Haoqian Wang. Nerf-ms: Neural radiance fields with multi-sequence. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 18591–18600, 2023. 3

[13] Yixuan Li, Lihan Jiang, Linning Xu, Yuanbo Xiangli, Zhenzhi Wang, Dahua Lin, and Bo Dai. Matrixcity: A large-scale city dataset for city-scale neural rendering and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3205–3215, 2023. 6

[14] Haotong Lin, Qianqian Wang, Ruojin Cai, Sida Peng, Hadar Averbuch-Elor, Xiaowei Zhou, and Noah Snavely. Neural scene chronology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20752–20761, 2023. 3, 6

[15] Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, et al. Vastgaussian: Vast 3d gaussians for large scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5166–5175, 2024. 3

[16] Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20654–20664, 2024. 2, 3, 5, 6, 8

[17] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *2024 International Conference on 3D Vision (3DV)*, pages 800–809. IEEE, 2024. 3

[18] Ricardo Martin-Brualla, David Gallup, and Steven M Seitz. 3d time-lapse reconstruction from internet photos. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1332–1340, 2015. 3

[19] Ricardo Martin-Brualla, David Gallup, and Steven M Seitz. Time-lapse mining from internet photos. *ACM Transactions on Graphics (TOG)*, 34(4):1–8, 2015. 3

[20] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7210–7219, 2021. 3

[21] Kevin Matzen and Noah Snavely. Scene chronology. In *European conference on computer vision*, pages 615–630. Springer, 2014. 3

[22] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2

[23] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 2

[24] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10318–10327, 2021. 2, 3

[25] Christian Reiser, Rick Szeliski, Dor Verbin, Pratul Srinivasan, Ben Mildenhall, Andreas Geiger, Jon Barron, and Peter Hedman. Merf: Memory-efficient radiance fields for real-time view synthesis in unbounded scenes. *ACM Transactions on Graphics (ToG)*, 42(4):1–12, 2023. 2

[26] Kerui Ren, Lihan Jiang, Tao Lu, Mulin Yu, Linning Xu, Zhangkai Ni, and Bo Dai. Octree-gs: Towards consistent real-time rendering with lod-structured 3d gaussians. *arXiv preprint arXiv:2403.17898*, 2024. 2

[27] Grant Schindler and Frank Dellaert. Probabilistic temporal inference on reconstructed 3d scenes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1410–1417. IEEE, 2010. 3

[28] Grant Schindler and Frank Dellaert. 4d cities: Analyzing, visualizing, and interacting with historical urban photo collections. *Journal of Multimedia*, 7(2), 2012. 3

[29] Grant Schindler, Frank Dellaert, and Sing Bing Kang. Inferring temporal order of images from 3d structure. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7. IEEE, 2007. 3

[30] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 6

[31] Mai Su, Zhongtao Wang, Huishan Au, Yilong Li, Xizhe Cao, Chengwei Pan, Yisong Chen, and Guoping Wang. Hug: Hierarchical urban gaussian splatting with block-based reconstruction for large-scale aerial scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 28839–28848, 2025. 2

[32] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 conference proceedings*, pages 1–12, 2023. 2

[33] Ali Osman Ulusoy and Joseph L Mundy. Image-based 4-d reconstruction using 3-d change detection. In *European Conference on Computer Vision*, pages 31–45. Springer, 2014. 3

[34] Yuze Wang, Junyi Wang, and Yue Qi. We-gs: An in-the-wild efficient 3d gaussian representation for unconstrained photo collections. *arXiv preprint arXiv:2406.02407*, 2024. 3

[35] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20310–20320, 2024. 3, 6

[36] Jiacong Xu, Yiqun Mei, and Vishal Patel. Wild-gs: Real-time novel view synthesis from unconstrained photo collections. *Advances in Neural Information Processing Systems*, 37:103334–103355, 2024. 3

[37] Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. *arXiv preprint arXiv:2310.10642*, 2023. 3, 6

[38] Lin Zeng, Boming Zhao, Jiarui Hu, Xujie Shen, Ziqiang Dang, Hujun Bao, and Zhaopeng Cui. Gaussianupdate: Continual 3d gaussian splatting update for changing environments. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 25800–25809, 2025. 3

[39] Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. In *European Conference on Computer Vision*, pages 341–359. Springer, 2024. 3, 6