SkyLume: A Large-Scale Multi-Illumination Aerial Benchmark for Urban Scene Reconstruction and Beyond

Zhuoxiao Li[1†], Wenzong Ma[1†], Taoyu Wu[2†], Jinjing Zhu[1,3], Shuai Zhang[1], Jing OU[1], Tongyan Hua[1], Yinrui Ren[1], Rongjun Qin[4], Hui Xiong[1*], and Wufan Zhao[1*]

[1] The Hong Kong University of Science and Technology (Guangzhou), China
[2] University of Liverpool, Liverpool, UK
[3] The Chinese University of Hong Kong, Hong Kong SAR, China
[4] The Ohio State University, Columbus, USA
[†] Equal contribution. [*] Corresponding authors.

Fig. 1: We present **SkyLume**, a comprehensive real-world UAV dataset which provides 6K-resolution five-directional imagery from three daily captures along identical RTK-guided flight paths, paired with LiDAR-derived ground truth including precise meshes and per-frame depth and normal maps. The benchmark enables rigorous evaluation of 3D reconstruction, novel view synthesis and inverse rendering at city-scale.

Abstract. Recent advances in Neural Radiance Fields and 3D Gaussian Splatting have demonstrated immense potential for large-scale UAV-based 3D reconstruction. However, real-world data collection often spans multiple times of the day, violating the fundamental photometric consistency assumption. Such multi-temporal illumination inconsistencies inevitably lead to severe color artifacts, coupled shadows, and degraded geometric accuracy. Due to the lack of UAV datasets that systematically capture the same areas under varying lighting conditions, this crucial challenge remains largely underexplored. To bridge this gap, we introduce

SkyLume, a large-scale, real-world UAV dataset specifically designed for studying illumination-robust 3D reconstruction and urban scene modeling. Our dataset features three primary contributions: (1) We collect over 100k high-resolution UAV images (nadir and four oblique views) spanning 10 diverse urban regions, with each region meticulously captured across three distinct time periods to systematically isolate illumination variations. (2) To enable rigorous evaluation, we provide comprehensive per-scene LiDAR scans, offering highly accurate ground truth for assessing depth, surface normals, and geometric fidelity. (3) For inverse rendering and appearance decoupling tasks, we introduce the Temporal Consistency Coefficient (TCC), a novel metric designed to quantify cross-time rendering stability. We envision **SkyLume** as a foundational benchmark that will advance research and real-world evaluation in large-scale inverse rendering, robust geometry reconstruction, and novel view synthesis. Project Page: skylume.

Keywords: Large-Scale 3D Reconstruction · Neural Radiance Fields · 3D Gaussian Splatting · Inverse Rendering · UAV Dataset

1 Introduction

UAV oblique photogrammetry has rapidly evolved into a standard tool for city-scale modeling [59]. While advancements in structure-from-motion (SfM) [32] and multi-view stereo (MVS) [33], Neural Radiance Fields (NeRF) [26], and 3D Gaussian Splatting (3DGS) [11] have significantly improved geometric completeness and rendering fidelity, multi-temporal data capture remains a critical bottleneck [2,3,19,22,25,36]. For city-scale mapping missions, data acquisition is often conducted in multiple batches over extended periods, with each area captured different illuminations. Current pipelines often bake these varying lighting conditions into textures or radiance fields, failing to maintain a consistent appearance when merging the results. [35, 49, 57]. This raises a central question: *Can current methods preserve geometric fidelity, photometric consistency, and novel-view realism under illumination shifts across different flight sessions?*

Answering this requires ***isolating the impact of illumination on the same scene***. Lighting variations drastically alter shadows, exposures, and multi-view correspondences, heavily biasing depth and normal estimations [13, 29, 35, 45]. However, existing UAV datasets fall short. They either rely on synthetic illumination from game engines [15, 58], or feature real-world captures that rarely revisit the same regions [9, 20, 39, 46]. Furthermore, they often lack the precise ground-truth geometry or multi-illumination conditions required to jointly evaluate appearance and structure under illumination shifts [23, 46]. Consequently, the challenge of illumination-robust 3D modeling remains underexplored due to the absence of a comprehensive **real-world** benchmark.

To bridge this gap, we present *SkyLume*, a large-scale, real-world UAV dataset tailored for illumination-robust urban 3D modeling. *SkyLume* encompasses 10 urban regions, each captured with five-direction imagery (nadir and four obliques) across three distinct times of day (morning, noon, evening). Crucially, all time slots are precisely aligned in a unified coordinate system and paired with per-scene LiDAR scans. We also provide per-frame ground-truth annotations for poses, depth, surface normals, and precise solar geometry from RTK time slot. Additionally, we introduce the Temporal Consistency Coefficient (TCC) to quantify cross-time albedo stability in inverse rendering. *SkyLume* transforms multi-temporal capture into a rigorous testbed: **methods must not only render observed views but also harmonize appearance and preserve underlying geometry across different daily illumination.**

Our extensive benchmark on representative 3DGS methods reveals significant vulnerabilities in current pipelines at scale: (1) **Rendering variance:** Illumination shifts drastically degrade rendering quality, causing divergent artifacts particularly on weakly textured façades and reflective surfaces. (2) **Geometric corruption:** Cast shadows and moving penumbras are frequently over-fitted as solid structures, introducing spurious geometry and biasing depth estimates. (3) **Fragile inverse rendering:** Estimated albedos often retain time-dependent shadows, failing to achieve true material decoupling. These findings underscore an urgent need for advanced paradigms that explicitly separate light from material and ensure robust cross-time consistency in large-scale reconstruction.

The main contributions of our work are summarized as follows:

- We present **SkyLume**, a large-scale, multi-temporal UAV oblique dataset specifically designed to drive research in illumination-robust 3D modeling. To the best of our knowledge, SkyLume is the first real-world aerial dataset comprising over 100K high-resolution images across 10 urban regions, rigorously paired with multi-time captures, high-precision LiDAR ground-truth geometry, and accurate 6-DoF poses.
- We establish standardized data splits and comprehensive evaluation protocols for cross-time rendering and geometric accuracy. Furthermore, we introduce the Temporal Consistency Coefficient (TCC) to quantitatively assess cross-time material decoupling in inverse rendering tasks.
- We systematically benchmark representative 3DGS variants under multi-illumination conditions. Our extensive evaluation exposes critical vulnerabilities in current pipelines regarding rendering fidelity, geometry extraction, and appearance consistency, thereby providing actionable insights and a rigorous testbed for future research.

2 Related Work

2.1 3D Gaussian Splatting and Beyond

3D Gaussian Splatting (3DGS) [11] represents a scene as an explicit set of anisotropic Gaussians and enables real-time, visibility-aware rendering with competitive fidelity. Since its introduction, a rapidly growing literature has extended 3DGS in terms of its applications, core algorithms, and compatibility:

Table 1: Comparison of aerial 3D reconstruction datasets. SkyLume is the only real-world dataset that supports rigorous evaluation of 3D reconstruction, NVS, and inverse rendering, while prior corpora miss one or more of these axes.

Aerial Dataset	Real-World	Lidar	Camera Type	Light	Depth/Normal	Resolution
ISPRS [27]	✓	✓ *(terrestrial)*	Oblique	✗	✗	6000×4000
UrbanScene3D [20]	✓ *(part)*	✓	Oblique	✗	✗	5490×3651
Mill 19 [36]	✓	✗	Oblique	✗	✗	4608×3456
Horizon-GS [9]	✓ *(part)*	✗	**Oblique+Nadir**	✗	✗	1600×1066
GauU-Scene [46]	✓	✓	**Oblique+Nadir**	✗	✗	5468×3636
UAVScenes [39]	✓	✓	Nadir	✗	*Depth*	N/A
Matrix City [15]	✗	✗ *(from depth)*	**Oblique+Nadir**	✓	**Depth+Normal**	1920×1080
SkyLume (ours)	✓	✓	**Oblique+Nadir**	✓	**Depth+Normal**	**6252×4168**

(a) Rendering-centric 3DGS aim for stable and real-time NVS. Key advances include anti-aliasing [52], improving densification [5, 12, 16, 24], improving visibility via deferred or reflective [45, 49, 57], appearance decoupling for view-dependent effects [45, 48], and LoD or region-aware scheduling to keep memory bounded [24, 30, 34]. On top of these foundations, large-scene variants mainly add partitioning and distributed optimization (with hierarchical/LoD training) [2, 17, 19, 21, 22, 54], boosting throughput while preserving fidelity.

(b) Geometry-aware 3DGS steers Gaussians toward surfaces to obtain editable, watertight geometry. Core ideas include surface-aligned Gaussians with fast mesh extraction and joint refinement [7], and surfel-like primitives that stabilize normals and coverage [1, 8, 43]. Hybrids with depth/normal-aware training [37, 53, 55] further improve mesh quality and downstream editability. **(c) Inverse-rendering 3DGS** augments Gaussians with explicit shading, materials, and illumination to enable relighting and editing. Representative directions include (i) joint recovery of geometry, BRDF, and environment lighting [18], and (ii) deferred/reflective pipelines that better handle speculars and visibility [10, 50, 57]; some further integrate differentiable ray tracing and environment Gaussians for real-time, view-dependent effects [45, 49].

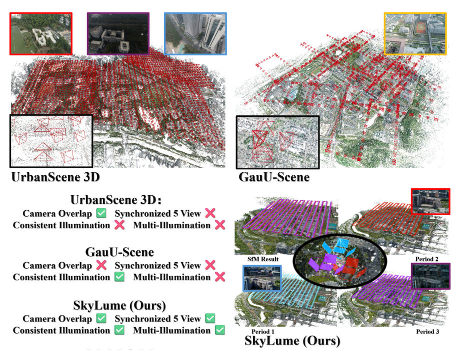

Fig. 2: Comparison with real-world UAV datasets. Skylume provides high overlap, 3 consistent multi-illumination and synchronized 5 view captures for every scene, while others missing consistent/multi illumination or high overlap.

Despite their strong performance, these methods are primarily designed for single-session captures and lack the ability to ensure multi-temporal consistency. This limitation is precisely the gap our benchmark addresses.

2.2 UAV Datasets for 3D Reconstruction

UAV datasets for 3D modeling fall into two buckets: **(a) Synthetic city-scale** corpora offer full ground truth, accurate poses, and controllable illumination/weather, which are ideal for ablations but carry domain gaps to real flights [6, 15, 38, 58]. These methods are built on game engines, provides aerial–street views with ground-truth cameras and flexible light/weather control for city-scale reconstruction [9, 31, 40]. **(b) Real-world captured** corpora better reflect deployment but rarely include cross-time revisits under controlled capture, complicating evaluation of illumination robustness. Some mix large synthetic cities with a limited set of real scenes, where capture conditions within a scene can vary [9, 23, 27, 36]. Some covers large urban areas with RGB and LiDAR and emphasizes reconstruction under observed lighting [46]. Some contributes frame-wise semantic labels for images and LiDAR, accurate 6-DoF poses, and reconstructed maps, broadening multi-modal evaluation but focusing primarily on perception rather than systematic multi-temporal oblique photogrammetry [4, 14, 28, 39, 44, 47].

As shown in Tab. 1 and Fig. 2, the aim of **SkyLume** is to address these above discussed gaps with a large-scale, real-world UAV *oblique* benchmark. It supports reproducible protocols for cross-time material/illumination consistency and façade-rich reconstruction at scale.

3 The SkyLume Dataset

The **Skylume** dataset is designed to address the limitations of existing datasets by providing comprehensive and high-quality data to support robust performance under varying illumination conditions. Fig. 3 illustrates the pipeline. Sec. 3.1 details the data acquisition setup, while Sec. 3.2 provides an overview of the data collection process. Sec. 3.3 outlines the data processing methods, and Sec. 3.4 presents the resulting high-quality data.

3.1 Equipment

As shown in Fig. 3, we employ a survey-grade setup designed for stable flights, high-resolution multi-view imagery, and accurate point clouds for reliable alignment and evaluation: *(a) DJI Matrice 350 RTK:* A UAV equipped with RTK-grade positioning and long endurance, enabling repeatable flights across different times of day. *(b) CHCNAV C30 Aerial Oblique Camera:* A 130 MP camera featuring four 45° oblique views and one 90° nadir view, captured synchronously to enhance façade coverage and minimize drift. *(c) DJI Zenmuse L2 LiDAR:* A frame-scanning LiDAR with 5 cm horizontal and 4 cm vertical accuracy at 150 m, ensuring precise alignment. *Detailed equipment specifications can be found in the supplementary material.*

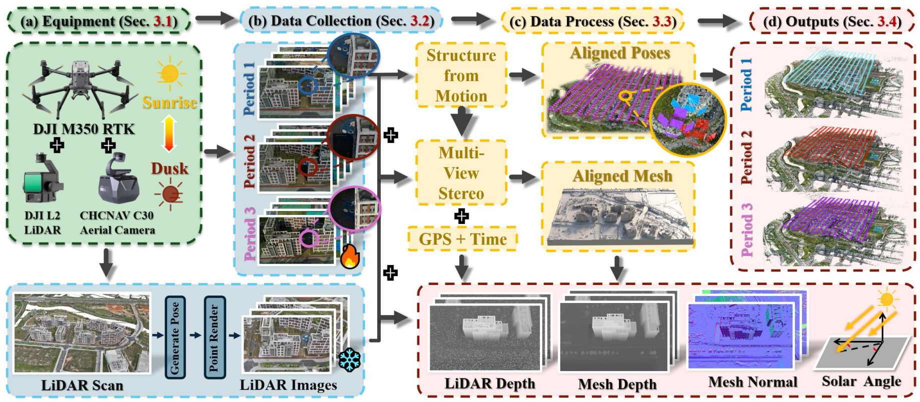

Fig. 3: Dataset collection and processing pipeline. (a) A survey-grade UAV stack flies the same RTK-guided route at three times of day to capture five-direction 6K imagery and LiDAR. (b) A unified LiDAR-guided SfM registration includes three periods and refines poses by point-rendering LiDAR into the cameras. (c) A LiDAR-guided MVS to produce a high-quality aligned ground-truth geometry. (d) We release per-period split SfM packages and per-frame LiDAR depth, mesh depth, mesh normals, and solar geometry.

3.2 Data Collection

To systematically isolate illumination variables, data collection is executed at three distinct periods of the day: early morning (07:00 – 09:00), midday (11:00 – 13:00), and late afternoon (16:00 – 18:00). Crucially, to ensure strict geometric consistency across varying lighting conditions, the exact same waypoint trajectory is automatically executed for each time slot, maintaining highly comparable spatial coverage and view directions.

The flight protocol strictly adheres to standard photogrammetric practices. The forward and side overlaps are set to 80% and 60%, respectively. Flights are conducted at an altitude of 120 m above ground level with an airspeed of 8 m/s. The CHCNAV C30 camera is triggered at 1 Hz, consistently capturing synchronized five-direction views. Real-Time Kinematic (RTK) corrections are continuously applied, and post-differential positioning is embedded in the EXIF metadata to secure highly accurate WGS 84 geo-location priors for all images. Since LiDAR is an active measurement modality unaffected by ambient light, it is acquired in a single dedicated flight on the following day to capture the definitive ground-truth geometry.

3.3 Data Preprocessing and Rigorous Alignment

Multi-Temporal SfM Alignment. The core challenge in our dataset construction is ensuring sub-pixel multi-modal registration. To achieve this, all data is projected into a unified coordinate system. The LiDAR point cloud is georeferenced in WGS 84 / UTM, while the RGB images rely on RTK-aided metadata.

As demonstrated in Fig. 3, We first register the LiDAR data and synthesize fixed camera poses to render pseudo-RGB anchor views from the point cloud. Subsequently, we execute an anchor-constrained joint SfM pipeline. Instead of independent reconstructions, all RGB images across the three time slots are jointly optimized alongside the LiDAR-derived **fixed** anchor views (see Fig. 4 (a)). The RTK poses serve as robust soft priors during the global bundle adjustment. In areas with sparse features where residual drift might persist, we augment the optimization with manually annotated Ground Control Points (GCPs), ensuring flawless cross-time and cross-modal registration

LiDAR-Guided Meshing. Leveraging the rigorously aligned multi-temporal poses, we perform dense surface reconstruction by fusing the three-period RGB imagery with the LiDAR data. The active LiDAR measurements act as a metric scaffold, significantly regularizing depth estimation in weakly textured or heavily shadowed regions where traditional MVS often fails.

Geometry Completion for Water Surfaces. Rivers and ponds inherently pose severe challenges for MVS due to view-dependent specularity and translucency. To rectify these topological artifacts, we leverage the LiDAR data to estimate physically plausible flat surfaces: For each water body, we sample four shoreline points to compute a mean elevation z_{mean} (ensuring a sample variance of $\leq$ 2 cm). The surface normals in these regions are explicitly constrained to be horizontal, and the patches are seamlessly re-meshed into the global geometry (see Fig. 4 (c)), guaranteeing structural continuity and high-quality ground truth.

3.4 Output Data

To ensure seamless integration with existing pipelines and facilitate reproducible evaluation, we export each reconstruction region in the COLMAP format.

SfM Results for Three Time Slots. From the global camera extrinsics, we isolate the *image name* corresponding to each of the three time slots, creating three distinct subsets of camera poses. Using these time-specific extrinsics, we then match and extract the corresponding camera intrinsics and sparse point cloud from the global SfM solution.

Table 2: The SfM alignment statistics and camera position uncertainty.

SfM alignment report	
Average Projections	13 921 072
Average track length	3.3
Maximum non-compressible error [px]	2.00
Median reprojection error [px]	0.70
Mean reprojection error [px]	0.79

Relative camera position uncertainty [m]			
	X	Y	Z
Mean	≤0.001	≤0.002	≤0.001
Standard deviation	≤0.001	≤0.001	≤0.001
Maximum	≤0.008	≤0.019	≤0.014
Minimum	≤0.001	≤0.001	≤0.001

Per-frame Depth and Normals. We provide two distinct modalities for depth and normal data: (a) *Mesh Depth and Normals* are derived by projecting the mesh into each camera view. The resulting depth and normal maps are dense, with resolutions matching those of the original images. (b) *LiDAR Depth* is obtained by re-projecting the LiDAR point cloud into the camera frames after occupancy filtering. Although this modality is sparser, it offers higher metric reliability for accurate depth estimation, as the removal of background objects prevents erroneous points from being projected.

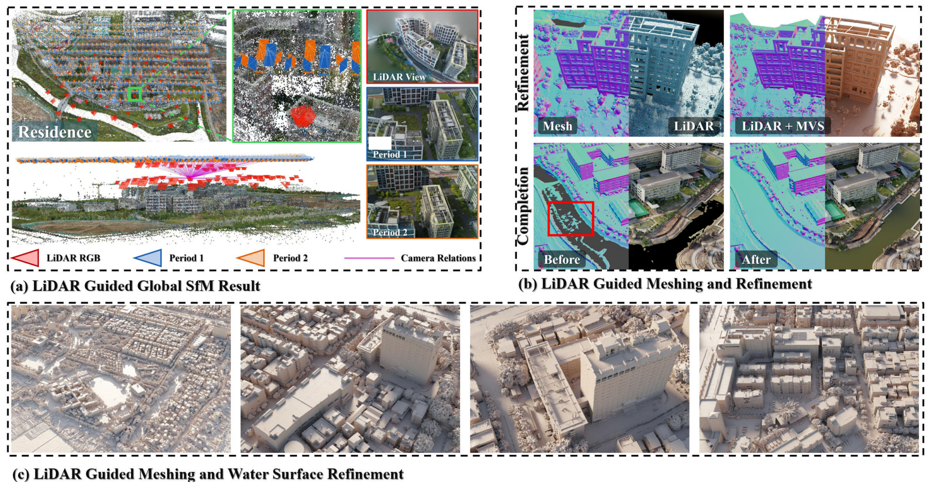

Fig. 4: Dataset Processing Demonstration. (a) We perform LiDAR-Guided SfM to register multi-temporal capture into a unified coordinate. We then split the unified SfM result and acquire three independent SfM result of each period; (b) We first build geometry ground truth via LiDAR-guided MVS. For reflective areas such as rivers and lakes, we manually repair water surfaces to correct MVS failures and ensure geometric continuity. (c) Final geometry ground-truth after LiDAR-guided refinement and water surface completion.

Table 3: Per-scene dataset statistics.

Scale	Scene Name	Total Images	Flight Height	Building Density	Illumination Type	Water Area	Glass Facade
Small Scale	**Gym**	6185	120.366	Low	Direct Sunlight (Morning) Direct Sunlight (Noon) Overcast (Afternoon)	No	No
	Staff Residence	7920	130.018	Medium	Direct Sunlight (Morning) Partly Cloudy (Noon) Overcast (Afternoon)	Yes	No
	iPark	5355	115.241	Medium	Direct Sunlight (Morning) Partly Cloudy (Noon) Overcast (Afternoon)	Yes	Yes
Medium Scale	**Tec School**	7185	108.481	Low	Direct Sunlight (Morning) Direct Sunlight (Noon) Direct Sunlight(Afternoon)	Yes	No
	Buildings	10455	129.891	High	Direct Sunlight (Morning) Direct Sunlight (Noon) Overcast (Afternoon)	No	Yes
	High School	10065	108.015	Medium	Direct Sunlight (Morning) Partly Cloudy (Noon) Direct Sunlight(Afternoon)	Yes	No
	Main Campus	10410	130.28	Medium	Direct Sunlight (Morning) Partly Cloudy (Noon) Direct Sunlight(Afternoon)	No	No
Large Scale	**Estate**	18630	109.024	High	Direct Sunlight (Morning) Partly Cloudy (Noon) Direct Sunlight(Afternoon)	No	No
	Town	12435	149.26	High	Direct Sunlight (Morning) Partly Cloudy (Noon) Overcast (Afternoon)	Yes	Yes
	Med School	20700	119.25	Medium	Direct Sunlight (Morning) Direct Sunlight (Noon) Overcast (Afternoon)	Yes	No

Per-frame solar geometry. Using the capture timestamp and geolocation of each image, we provide solar elevation and solar azimuth using the NOAA Solar Geometry Calculator. We hope these annotations can support future deshadowing, inverse rendering, and relighting studies.

Applications. Given the scale, rigorous alignment precision, and rich attribute diversity, our dataset serves as a versatile testbed for a wide array of downstream computer vision tasks. It is ideally suited for benchmarking large-scale illumination-robust novel view synthesis, urban reconstruction, and complex material estimation. Crucially, by systematically isolating illumination variables and providing high-fidelity LiDAR ground truth, our dataset introduces a unique benchmark for evaluating the photometric robustness of classical Structure-from-Motion (SfM) and Multi-View Stereo (MVS) pipelines under dynamic lighting. Furthermore, the availability of precise metric depth and extensive multi-view aerial coverage unlocks new opportunities to train and validate aerial monocular depth estimation networks, as well as emerging large-scale feed-forward 3D reconstruction models, thereby driving the next generation of generalizable aerial vision.

4 Benchmark Experiments

In this section, we evaluate three tracks on six small- and medium-scale scenes: inverse rendering methods GS-IR [18], Ref-Gaussian [49], and Ref-GS [57]; geometry methods 2DGS [8], PGSR [1], GOF [53], and CitygaussianV2 [22]; and novel view synthesis methods 3DGS [11], Abs-GS [51], Mip-Splatting [52], and Octree-GS [30]. To better evaluate the performance of each method in large-scale scenes, we conducted benchmarks on an NVIDIA A800 80G GPU, applying identical adjustments to the training pipeline and parameter settings for each method. *For detailed implementation information, please refer to the supplementary material.*

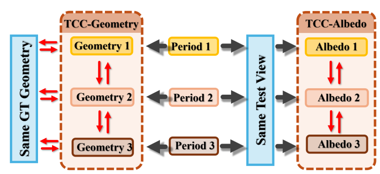

Fig. 5: Temporal Consistency Evaluation. For inverse rendering, we render albedo from identical test viewpoints across different time periods. For geometry, we evaluate each mesh against the ground truth and compute temporal consistency by averaging the pairwise consistency between period-specific meshes.

4.1 Benchmark Metrics

Multi-Temporal Illumination Robustness. In real-world large-scale scenarios, acquiring accurate ground-truth albedo is highly infeasible. To evaluate the stability and material decoupling ability of inverse rendering methods under varying lighting, we propose the Temporal Consistency Coefficient (**TCC**) framework, as illustrated in Fig. 5. Since our dataset guarantees that all images across different times of day are rigorously registered within a unified SfM coordinate system, we can synthesize albedo from identical virtual viewpoints across models trained independently on different time slots.

Specifically, from the global trajectory, we sample K fixed test viewpoints $\{v_k\}_{k=1}^{K}$ ensuring uniform spatial coverage. For each viewpoint v_k, we evaluate methods trained independently on $T = 3$ distinct time slots. Let $A_t^{(k)} \in$

$[0,1]^{H\times W\times 3}$ denote the rendered albedo from the model trained on time slot $t \in \{1,2,3\}$, and let $\bar{A}^{(k)} = \frac{1}{T}\sum_{t=1}^{T} A_t^{(k)}$ be the cross-time mean albedo.

Rather than relying on a singular combined metric, the TCC framework quantifies stability across multi-dimensional criteria. We compute the deviation of each time-specific albedo $A_t^{(k)}$ against the mean albedo $\bar{A}^{(k)}$. The structure preservation is evaluated via $\mathrm{TCC}_{\mathrm{SSIM}}$, and the perceptual consistency via $\mathrm{TCC}_{\mathrm{LPIPS}}$, and we formulate $\mathrm{TCC}_{\mathrm{MAE}}$ and $\mathrm{TCC}_{\mathrm{RMSE}}$ to measure pixel-level color stability:

$$\mathrm{TCC}_{\mathrm{S,L,M,R}} = \frac{1}{K\cdot T}\sum_{k=1}^{K}\sum_{t=1}^{T}\mathrm{SSIM,LPIPS,MAE,RMSE}\big(A_t^{(k)},\bar{A}^{(k)}\big). \tag{1}$$

A robust inverse rendering method should effectively decouple materials from illumination, yielding high $\mathrm{TCC}_{\mathrm{SSIM}}$ scores while minimizing $\mathrm{TCC}_{\mathrm{MAE}}$, $\mathrm{TCC}_{\mathrm{RMSE}}$, and $\mathrm{TCC}_{\mathrm{LPIPS}}$ errors.

To provide a unified ranking for leaderboard comparisons, we further aggregate these four-dimensional metrics into a single Overall Temporal Consistency Coefficient ($\mathrm{TCC}_{\mathrm{overall}} \in [0,1]$), where a higher value indicates stronger cross-time stability. Since the rendered albedo values are normalized to $[0,1]$, the error metrics ($\mathrm{TCC}_{\mathrm{MAE}}$ and $\mathrm{TCC}_{\mathrm{RMSE}}$) and the perceptual distance ($\mathrm{TCC}_{\mathrm{LPIPS}}$) inherently fall within the same range. Thus, we invert these error-based metrics into similarity scores and compute the arithmetic mean alongside $\mathrm{TCC}_{\mathrm{SSIM}}$:

$$\mathrm{TCC} = \frac{1}{4}\Big((1-\mathrm{TCC}_{\mathrm{MAE}})+(1-\mathrm{TCC}_{\mathrm{RMSE}})+\mathrm{TCC}_{\mathrm{SSIM}}+(1-\mathrm{TCC}_{\mathrm{LPIPS}})\Big). \tag{2}$$

A perfect inverse rendering method that achieves absolute material decoupling across varying illumination will yield a $\mathrm{TCC}_{\mathrm{overall}}$ score of exactly 1.

Table 4: TCC metric across three inverse-rendering baselines. We report the distribution of sub-metrics and the combined **TCC Overall** with Mean, Min, Max, and Std. All scores lie in [0,1], higher is better except Std.

	Mean↑			**Min↑**			**Max↑**			**Std↓**		
methods	GS-IR	Ref-Gaussian	Ref-GS	GS-IR	Ref-Gaussian	Ref-GS	GS-IR	Ref-Gaussian	Ref-GS	GS-IR	Ref-Gaussian	Ref-GS
TCC-LPIPS	0.826	0.866	**0.874**	0.759	**0.778**	0.755	0.867	0.919	**0.979**	**0.020**	0.027	0.029
TCC-SSIM	0.905	0.883	**0.928**	0.864	0.832	**0.878**	0.936	0.928	**0.985**	**0.014**	0.019	0.017
TCC-MAE	0.700	0.513	**0.766**	0.212	0.262	**0.587**	0.800	0.653	**0.977**	0.074	0.078	**0.063**
TCC Overall	0.721	0.658	**0.775**	0.563	0.575	**0.622**	0.765	0.735	**0.913**	0.033	**0.026**	0.036

Geometry robustness. Demonstrated in Fig. 5, we evaluate robustness of reconstructed geometry under changing illumination in two ways. First, we measure absolute accuracy against per scene GT geometry. Second, we measure cross-time consistency between meshes reconstructed from the three time slots.
Novel view synthesis. We additionally evaluate novel view synthesis using standard image quality metrics, reporting PSNR, SSIM [41], and LPIPS [56].

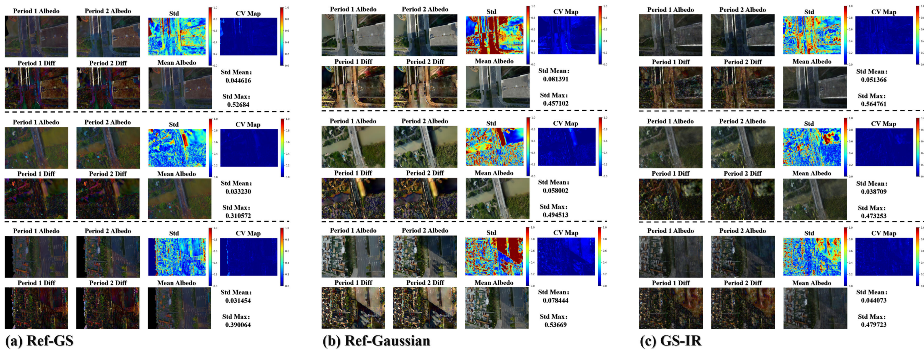

Fig. 6: Albedo and TCC visualization across two time slots. Note that *Period 2* is rendered from the viewpoints of *Period 1* to fix camera pose, so differences arise solely from illumination.

4.2 Cross-Time Albedo Evaluation

We assess temporal albedo consistency across three inverse-rendering baselines using our TCC metric, which rewards cross-time stability rather than single-slot fidelity. Tab. 4 shows that albedo estimates drift under illumination changes even at matched viewpoints, with large dispersion across scenes in the Min, Max, and Std columns. The perceptual branch TCC-MAE is frequently the limiting factor while TCC-SSIM remains comparatively stable, indicating that **structure is preserved more often than appearance**. Fig. 6 corroborates these trends on two regions with strong lighting shifts where all methods retain shadow imprinting. As a result, the proposed TCC turns multi-temporal robustness into a measurable target, enabling fair and reproducible comparison of illumination-robust inverse rendering.

Table 5: TCC metric across four geometry baselines. For each method we report precision (Pre), recall (Rec), and F-1 at three distance tolerances $\tau \in \{0.25, 0.5, 0.75\}$ m for Period 0 and Period 1. The rightmost block provides temporal consistency by averaging the pairwise F-1 scores across period meshes at the same distance tolerances.

	Period 1									**Period 2**									**TCC-Geometry**		
τ (meters)	0.25			0.5			0.75			0.25			0.5			0.75			0.25	0.5	0.75
metrics	Pre↑	Rec↑	F-1 ↑	Pre↑	Rec↑	F-1	Pre↑	Rec↑	F-1↑	Pre↑	Rec↑	F-1	Pre↑	Rec↑	F-1 ↑	Pre↑	Rec↑	F-1↑	F-1↑	F-1↑	F-1↑
2DGS [8]	0.203	0.259	0.227	0.462	0.562	0.508	0.663	0.750	0.704	0.199	**0.278**	0.232	0.486	**0.639**	0.552	0.672	**0.811**	0.735	**0.570**	0.675	0.750
PGSR [1]	0.136	0.171	0.152	0.382	0.485	0.428	0.680	0.764	0.719	0.150	0.198	0.171	0.368	0.489	0.420	0.671	0.778	0.720	0.554	0.662	0.749
GOF [53]	0.106	0.101	0.103	0.417	0.336	0.372	0.775	0.594	0.672	0.120	0.113	0.116	0.455	0.357	0.400	**0.810**	0.599	0.689	0.431	0.610	0.712
CityGaussianV2 [22]	**0.315**	**0.270**	**0.291**	**0.618**	**0.603**	**0.610**	**0.782**	**0.835**	**0.808**	**0.256**	0.245	**0.250**	**0.563**	0.569	**0.566**	0.797	0.746	**0.770**	0.547	**0.719**	**0.790**

4.3 Cross-Time Geometry Evaluation

We evaluate geometry under illumination change by reporting precision, recall, and F-1 against GT for each time slot, and by measuring cross-time consistency

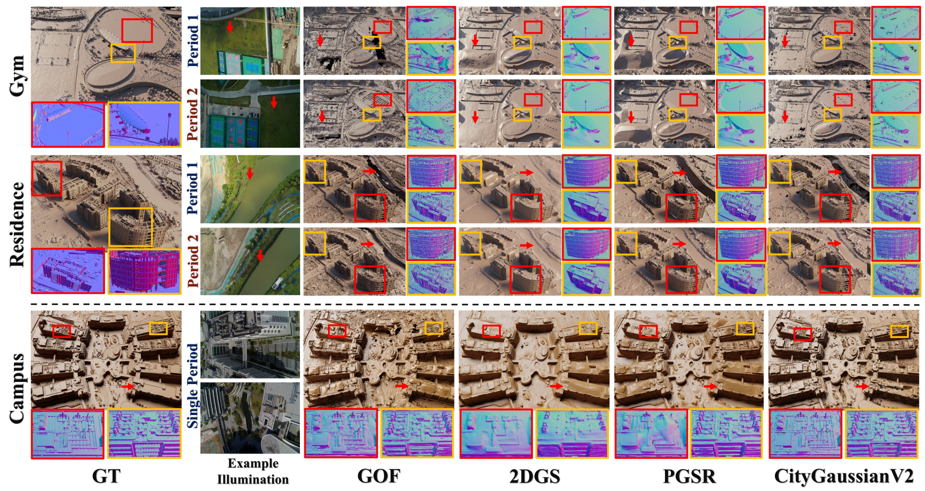

Fig. 7: Geometry visualization. Top: Geometry for *Gym* and *Residence*, red and yellow frames are the zoomed-in surface normals. **Bottom**: Single Period 1 mesh comparison. Under the sunlit *Period 1*, all methods exhibit holes and breakups.

via the average pairwise consistency F-1 score between the meshes in Tab. Tab. 5. Two lighting regimes are highlighted in Fig. 7: a sunlit case with strong, moving shadows and color shifts, and an overcast case dominated by diffuse irradiance. Geometry extracted in diffuse conditions is the most stable, while direct sunlight degrades both absolute accuracy and cross-time agreement.

Methods that use alpha-blending depth for TSDF-fusion [42] (2DGS [8] and PGSR [1]) preserve mesh topology more consistently across slots but lack geometry details. Approaches that rely on opacity (GOF [53] and CityGaussianV2 [22]) can reach high per-slot F-1 score under favorable lighting but are brittle under hard shadows. Fig. 7 reveals holes and breakups in sunlight-swept regions such as lawn and water-body.

4.4 Novel View Synthesis Evaluation

Under strong lighting conditions, we evaluate four NVS-oriented methods and three geometry-oriented methods. As summarized in Tab. Tab. 6, the trend is consistent across scenes: geometry-oriented pipelines incur a noticeable loss in photometric fidelity and therefore lag in rendering quality. Within the NVS group, Abs-GS [51] is the most reliable under UAV oblique viewpoints, and Octree-GS [30] recovers fine structures. By contrast, 2DGS [8] produces blurred renderings in all four representative scenes, as illustrated in Fig. 8.

4.5 Other Tasks to Explore

Beyond our primary benchmark, SkyLume offers a rich testbed for exploring diverse 3D vision tasks. As illustrated in Fig. 9, we conduct pilot evaluations

Table 6: Novel view synthesis under *Period 1* (sunlit) reconstruction. We report SSIM, PSNR, and LPIPS across six scenes for four NVS-oriented methods and three geometry-capable pipelines. "-" indicates no results due to GPU out-of-memory.

	GYM			iPark			Tec School			Staff residence			Campus			High School		
metrics	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
3DGS	0.672	21.144	0.318	0.707	**23.250**	0.355	0.631	21.652	0.384	0.670	23.099	0.353	0.617	**21.507**	0.409	0.626	20.570	0.389
Abs-GS	**0.681**	**21.183**	**0.305**	**0.711**	22.958	**0.335**	**0.639**	21.523	0.375	**0.676**	**23.125**	**0.352**	**0.633**	21.477	**0.390**	**0.634**	20.413	**0.382**
Mip-Splatting	0.669	21.131	0.325	0.701	22.708	0.362	0.621	21.364	0.398	0.667	23.006	0.360	0.619	20.015	0.427	-	-	-
Octree-GS	0.648	20.931	0.336	0.697	22.681	0.358	0.607	**21.765**	**0.360**	0.659	23.094	0.366	0.624	21.432	0.425	0.624	**20.618**	**0.380**
GOF	0.626	20.830	0.360	0.681	23.172	0.382	0.619	21.224	0.401	0.619	22.513	0.401	0.505	20.047	0.524	-	-	-
2DGS	0.583	20.350	0.425	0.648	22.722	0.440	0.549	21.103	0.489	0.585	21.891	0.458	0.507	20.518	0.539	0.497	19.531	0.539
PGSR	0.580	20.110	0.424	0.622	21.953	0.467	0.537	20.926	0.500	0.591	22.094	0.426	0.514	20.572	0.530	0.490	19.398	0.547

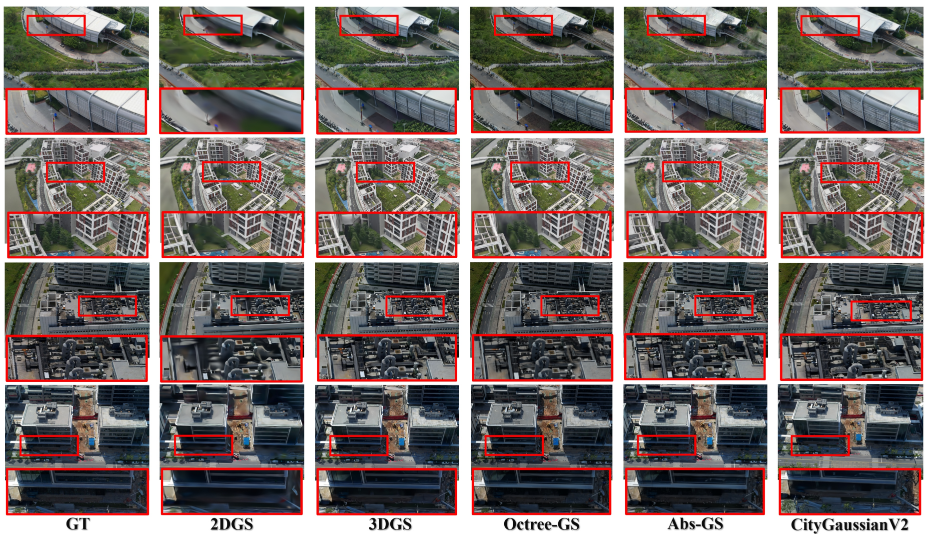

Fig. 8: Visualization of novel view synthesis under *Period 1* (sunlit). We visualize shadow-dominated regions to highlight some baselines lost details.

on several emerging paradigms to highlight the dataset's broader utility and the outstanding challenges in aerial vision.

Monocular Depth and Normal Estimation. Evaluating state-of-the-art monocular depth and normal estimators reveals that while current models maintain general stability and capture the macroscopic layout from UAV perspectives, they severely lack fine-grained structural details. The predicted geometries are often overly smoothed, failing to represent sharp architectural edges or intricate roof topologies. Our high-fidelity LiDAR ground truth provides a much-needed supervision signal and evaluation metric to drive the development of high-frequency, aerial-specific monocular estimators.

Multi-View Stereo (MVS) under Dynamic Lighting. We evaluate classical dense reconstruction pipelines (e.g., COLMAP) across our multi-temporal splits. The results expose severe vulnerabilities to dynamic illumination: while midday models are relatively complete, conditions with low ambient exposure or elongated cast shadows consistently break the multi-view photometric consis-

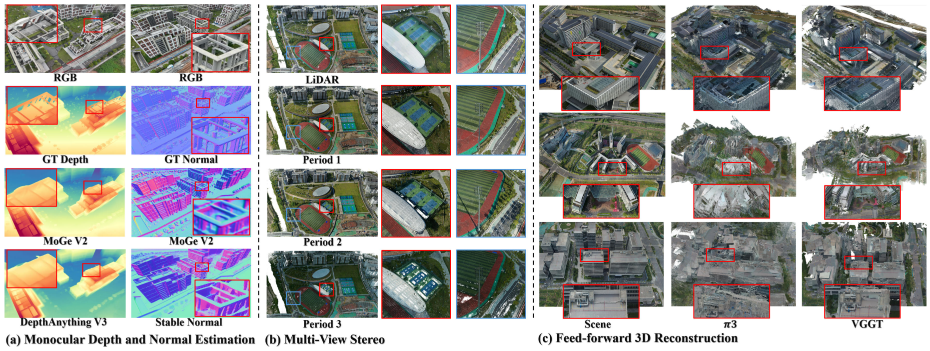

Fig. 9: Other Tasks to Explore.

tency assumption. This leads to massive structural holes, disconnected surfaces, and severe noise in the resulting dense point clouds, underscoring the urgent need for illumination-robust feature matching and geometry extraction algorithms.

Feed-forward 3D Reconstruction. We also benchmark recent large-scale feed-forward reconstruction networks (such as π^3 and vggt) directly on our raw, distorted UAV imagery. Observations indicate that these single-forward-pass models remain highly unstable under unconstrained aerial trajectories. Specifically, π^3 exhibits massive geometric overlapping artifacts, indicating severe failures in joint intrinsic and extrinsic camera pose estimation from raw distorted images. While vggt demonstrates slightly better resilience to these distortions, the overall geometric fidelity remains insufficient for city-scale modeling. This highlights a significant domain gap between object-centric feed-forward models and large-scale, raw aerial captures.

5 Conclusion and Future Work

In this paper, we introduce **SkyLume**, a large-scale UAV dataset that couples high-resolution imagery captured under three different illumination conditions. The release includes unified SfM resolution, per-frame depth and normals, and solar geometry. Central to the dataset is the Temporal Consistency Coefficient metric, which evaluates cross-time stability rather than single-slot fidelity. Experiments findings point to concrete directions: future research should integrate cross-time, shadow-aware priors, and jointly reason about illumination, material, and geometry at city scale.

Future Work. We plan to further expand SkyLume with additional scenes, weather conditions, and sensing modalities to enhance robustness and support tasks such as relighting and editing. In addition, we plan to add multimodal annotations, such as semantic segmentation and infrared imagery, to enable cross-modal fusion for illumination-robust reconstruction. Our goal is to transform multi-temporal UAV capture from a confounding factor into a powerful tool for stable inverse rendering, reliable geometry reconstruction, and high-fidelity novel view synthesis in real urban environments.

Acknowledgments

This work was supported in part by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 42401567), Guangdong Provincial Project (Grant No. 2024QN11G095), The Open Fund of the State Key Laboratory of Spatial Datum (Grant No. SKLSD2026-KF-25), National Key R&D Program of China (Grant No. 2023YFF0725001), in part by the National Natural Science Foundation of China (Grant No. 92370204), in part by the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2023B1515120057), in part by the Key-Area Special Project of Guangdong Provincial Ordinary Universities (Grant No. 2024ZDZX1007), in part by the Guangdong Provincial Project (Grant No. 2024QN11G095), and in part by the Research Grants Council (RGC) of Hong Kong SAR (Grant Nos. GRF14213125, GRF14201824, and GRF14216222).

References

1. Chen, D., Li, H., Ye, W., Wang, Y., Xie, W., Zhai, S., Wang, N., Liu, H., Bao, H., Zhang, G.: Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction (2024)
2. Chen, J., Ye, W., Wang, Y., Chen, D., Huang, D., Ouyang, W., Zhang, G., Qiao, Y., He, T.: Gigags: Scaling up planar-based 3d gaussians for large scene surface reconstruction (2024), https://arxiv.org/abs/2409.06685
3. Chen, Y., Lee, G.H.: Dogaussian: Distributed-oriented gaussian splatting for large-scale 3d reconstruction via gaussian consensus (2024), https://arxiv.org/abs/2405.13943
4. Dhrafani, D., Liu, Y., Jong, A., Shin, U., He, Y., Harp, T., Hu, Y., Oh, J., Scherer, S.: Firestereo: Forest infrared stereo dataset for uas depth perception in visually degraded environments. arXiv preprint arXiv:2409.07715 (2024)
5. Fang, G., Wang, B.: Mini-splatting: Representing scenes with a constrained number of gaussians (2024), https://arxiv.org/abs/2403.14166
6. Fonder, M., Van Droogenbroeck, M.: Mid-air: A multi-modal dataset for extremely low altitude drone flights. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0 (2019)
7. Guédon, A., Lepetit, V.: Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. arXiv preprint arXiv:2311.12775 (2023)
8. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: SIGGRAPH 2024 Conference Papers. Association for Computing Machinery (2024). https://doi.org/10.1145/3641519.3657428
9. Jiang, L., Ren, K., Yu, M., Xu, L., Dong, J., Lu, T., Zhao, F., Lin, D., Dai, B.: Horizon-gs: Unified 3d gaussian splatting for large-scale aerial-to-ground scenes. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 26789–26799 (2025)

10. Jiang, Y., Tu, J., Liu, Y., Gao, X., Long, X., Wang, W., Ma, Y.: Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5322–5332 (2024)
11. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
12. Kheradmand, S., Rebain, D., Sharma, G., Sun, W., Tseng, Y.C., Isack, H., Kar, A., Tagliasacchi, A., Yi, K.M.: 3d gaussian splatting as markov chain monte carlo. In: Advances in Neural Information Processing Systems (NeurIPS) (2024), spotlight Presentation
13. Langguth, F., Sunkavalli, K., Hadap, S., Goesele, M.: Shading-aware multi-view stereo. In: European Conference on Computer Vision. pp. 469–485. Springer (2016)
14. Li, W., Yu, S., Wang, C., Hu, G., Shen, S., Wen, C.: SGLoc: Scene geometry encoding for outdoor LiDAR localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9286–9295 (2023)
15. Li, Y., Jiang, L., Xu, L., Xiangli, Y., Wang, Z., Lin, D., Dai, B.: Matrixcity: A large-scale city dataset for city-scale neural rendering and beyond. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3205–3215 (2023)
16. Li, Z., Yao, S., Chu, Y., Garcia-Fernandez, A.F., Yue, Y., Lim, E.G., Zhu, X.: Mvg-splatting: Multi-view guided gaussian splatting with adaptive quantile-based geometric consistency densification (2024), `https://arxiv.org/abs/2407.11840`
17. Li, Z., Yao, S., Wu, T., Yue, Y., Zhao, W., Qin, R., Ángel F. García-Fernández, Levers, A., Ralph, J., Zhu, X.: Ulsr-gs: Urban large-scale surface reconstruction gaussian splatting with multi-view geometric consistency. ISPRS Journal of Photogrammetry and Remote Sensing **230**, 861–880 (2025). `https://doi.org/https://doi.org/10.1016/j.isprsjprs.2025.10.008`, `https://www.sciencedirect.com/science/article/pii/S092427162500396X`
18. Liang, Z., Zhang, Q., Feng, Y., Shan, Y., Jia, K.: Gs-ir: 3d gaussian splatting for inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21644–21653 (2024)
19. Lin, J., Li, Z., Tang, X., Liu, J., Liu, S., Liu, J., Lu, Y., Wu, X., Xu, S., Yan, Y., Yang, W.: Vastgaussian: Vast 3d gaussians for large scene reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5166–5175 (June 2024)
20. Lin, L., Liu, Y., Hu, Y., Yan, X., Xie, K., Huang, H.: Capturing, reconstructing, and simulating: the urbanscene3d dataset. In: European Conference on Computer Vision. pp. 93–109. Springer (2022)
21. Liu, Y., Guan, H., Luo, C., Fan, L., Wang, N., Peng, J., Zhang, Z.: Citygaussian: Real-time high-quality large-scale scene rendering with gaussians. arXiv preprint arXiv:2404.01133 (2024)
22. Liu, Y., Luo, C., Mao, Z., Peng, J., Zhang, Z.: Citygaussianv2: Efficient and geometrically accurate reconstruction for large-scale scenes. In: ICLR (2025)
23. Liu, Y., Xue, F., Huang, H.: Urbanscene3d: A large scale urban scene dataset and simulator (2021)
24. Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., Dai, B.: Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20654–20664 (2024)
25. Mi, Z., Xu, D.: Switch-nerf: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In: International Conference on Learning Representations (ICLR) (2023), `https://openreview.net/forum?id=PQ2zoIZqvm`

26. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
27. Nex, F., Gerke, M., Remondino, F., Przybilla, H.J., Bäumker, M., Zurhorst, A.: Isprs benchmark for multi-platform photogrammetry. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **II-3/W4**, 135–142 (2015). https://doi.org/10.5194/isprsannals-II-3-W4-135-2015, https://isprs-annals.copernicus.org/articles/II-3-W4/135/2015/
28. Nguyen, T.M., Yuan, S., Cao, M., Lyu, Y., Nguyen, T.H., Xie, L.: Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint. The International Journal of Robotics Research **41**(3), 270–280 (2022)
29. Philip, J., Gharbi, M., Zhou, T., Efros, A.A., Drettakis, G.: Multi-view relighting using a geometry-aware network. ACM Trans. Graph. **38**(4), 78–1 (2019)
30. Ren, K., Jiang, L., Lu, T., Yu, M., Xu, L., Ni, Z., Dai, B.: Octree-gs: Towards consistent real-time rendering with lod-structured 3d gaussians. arXiv preprint arXiv:2403.17898 (2024)
31. Rizzoli, G., Barbato, F., Caligiuri, M., Zanuttigh, P.: Syndrone-multi-modal uav dataset for urban scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2210–2220 (2023)
32. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4104–4113 (2016)
33. Seitz, S.M., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). vol. 1, pp. 519–528. IEEE (2006)
34. Seo, Y., Choi, Y.S., Son, H.S., Uh, Y.: Flod: Integrating flexible level of detail into 3d gaussian splatting for customizable rendering (2024), https://arxiv.org/abs/2408.12894
35. Song, S., Qin, R.: A general albedo recovery approach for aerial photogrammetric images through inverse rendering. ISPRS Journal of Photogrammetry and Remote Sensing **218**, 101–119 (2024). https://doi.org/https://doi.org/10.1016/j.isprsjprs.2024.09.001, https://www.sciencedirect.com/science/article/pii/S0924271624003319
36. Turki, H., Ramanan, D., Satyanarayanan, M.: Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12922–12931 (June 2022)
37. Turkulainen, M., Ren, X., Melekhov, I., Seiskari, O., Rahtu, E., Kannala, J.: Dn-splatter: Depth and normal priors for gaussian splatting and meshing (2024)
38. Vuong, K., Ghosh, A., Ramanan, D., Narasimhan, S., Tulsiani, S.: Aerialmegadepth: Learning aerial-ground reconstruction and view synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2025)
39. Wang, S., Li, S., Zhang, Y., Yu, S., Yuan, S., She, R., Guo, Q., Zheng, J., Howe, O.K., Chandra, L., et al.: Uavscenes: A multi-modal dataset for uavs. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 28946–28958 (2025)
40. Wang, W., Zhu, D., Wang, X., Hu, Y., Qiu, Y., Wang, C., Hu, Y., Kapoor, A., Scherer, S.: Tartanair: A dataset to push the limits of visual slam (2020)

41. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
42. Werner, D., Al-Hamadi, A., Werner, P.: Truncated signed distance function: experiments on voxel size. In: Image Analysis and Recognition: 11th International Conference, ICIAR 2014, Vilamoura, Portugal, October 22-24, 2014, Proceedings, Part II 11. pp. 357–364. Springer (2014)
43. Wolf, Y., Bracha, A., Kimmel, R.: Surface reconstruction from gaussian splatting via novel stereo views. arXiv preprint arXiv:2404.01810 (2024)
44. Wu, R., Cheng, X., Zhu, J., Liu, Y., Zhang, M., Yan, S.: Uavd4l: A large-scale dataset for uav 6-dof localization. In: 2024 International Conference on 3D Vision (3DV). pp. 1574–1583. IEEE (2024)
45. Xie, T., Chen, X., Xu, Z., Xie, Y., Jin, Y., Shen, Y., Peng, S., Bao, H., Zhou, X.: Envgs: Modeling view-dependent appearance with environment gaussian. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5742–5751 (2025)
46. Xiong, B., Zheng, N., Liu, J., Li, Z.: Gauu-scene v2: Assessing the reliability of image-based metrics with expansive lidar image dataset using 3dgs and nerf. arXiv preprint arXiv:2404.04880 (2024)
47. Yan, Q., Zheng, J., Reding, S., Li, S., Doytchinov, I.: Crossloc: Scalable aerial localization assisted by multimodal synthetic data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17358–17368 (2022)
48. Yang, Z., Gao, X., Sun, Y., Huang, Y., Lyu, X., Zhou, W., Jiao, S., Qi, X., Jin, X.: Spec-gaussian: Anisotropic view-dependent appearance for 3d gaussian splatting (2024), `https://arxiv.org/abs/2402.15870`
49. Yao, Y., Zeng, Z., Gu, C., Zhu, X., Zhang, L.: Reflective gaussian splatting. arXiv preprint arXiv:2412.19282 (2024)
50. Ye, K., Hou, Q., Zhou, K.: 3d gaussian splatting with deferred reflection. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–10 (2024)
51. Ye, Z., Li, W., Liu, S., Qiao, P., Dou, Y.: Absgs: Recovering fine details for 3d gaussian splatting (2024)
52. Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-splatting: Alias-free 3d gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19447–19456 (2024)
53. Yu, Z., Sattler, T., Geiger, A.: Gaussian opacity fields: Efficient high-quality compact surface reconstruction in unbounded scenes. arXiv:2404.10772 (2024)
54. Yu Chen, G.H.L.: Dogaussian: Distributed-oriented gaussian splatting for large-scale 3d reconstruction via gaussian consensus. In: arXiv (2024)
55. Zhang, B., Fang, C., Shrestha, R., Liang, Y., Long, X., Tan, P.: Rade-gs: Rasterizing depth in gaussian splatting. arXiv preprint arXiv:2406.01467 (2024)
56. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
57. Zhang, Y., Chen, A., Wan, Y., Song, Z., Yu, J., Luo, Y., Yang, W.: Ref-gs: Directional factorization for 2d gaussian splatting. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 26483–26492 (2025)
58. Zhao, F., Guo, Y., Yang, S., Chen, X., Wang, L., Xu, L., Zhang, Y., Shi, Y., Yu, J.: Aerialgo: Walking-through city view generation from aerial perspectives. arXiv preprint arXiv:2412.00157 (2024)

59. Zhou, T., Lv, L., Liu, J., Wan, J.: Application of uav oblique photography in real scene 3d modeling. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **43**, 413–418 (2021)