GAME2GS: COVERAGE-AWARE GAUSSIANIZATION FOR LARGE-SCALE FREE-VIEW HDR RENDERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Modern game worlds combine detailed geometry, materials, lighting, and engine-specific rendering effects, making their visual appearance costly to transfer across platforms. Distilling this appearance into 3D Gaussian representations offers a way to reuse game assets beyond the source engine, but requires faithful 6DoF rendering and exposure adaptation under complex lighting. Existing large-scale reconstruction pipelines are typically trained on images captured along prescribed street trajectories. Denser sampling along these trajectories increases acquisition costs without necessarily providing sufficiently resolved and directionally diverse observations of all relevant surfaces. This motivates selecting more informative views under a limited image budget. These views must also capture the world's bright and dark regions on a consistent radiometric scale to support runtime exposure adaptation. However, auto-exposed LDR capture introduces inconsistent brightness supervision across views, whereas fixed-exposure LDR capture can lose highlight or shadow detail. We introduce Game2GS, a framework combining coverage-aware view acquisition with HDR Gaussian reconstruction. Using known source geometry, we select camera poses derived from baked lighting probes to improve surface coverage through well-resolved observations from complementary directions and distinct positions. We then reconstruct an HDR Gaussian scene from fixed-exposure captures and enable real-time automatic exposure using an engine-exported compensation curve. Under matched image budgets, Game2GS improves PSNR by more than 3 dB over the strongest prescribed-view baseline across all evaluated scenes on independent 6DoF test views, and we will release our private game scenes.

1 INTRODUCTION

Recent advances in 3D Gaussian Splatting (3DGS) have enabled high-quality novel-view synthesis with real-time rendering, offering a promising appearance-centric representation for complex environments (Kerbl et al., 2023). Modern AAA game worlds derive their visual fidelity from detailed geometry, rich materials and textures, and sophisticated lighting and shading (Akenine-Möller et al., 2018). Deploying such content across heterogeneous platforms, particularly on resource-constrained devices, often requires substantial asset simplification and artist-driven re-authoring (Gregory, 2018; Liang et al., 2023). Meanwhile, a growing body of work has demonstrated that Gaussian representations can be compressed, organized into levels of detail, and streamed according to available rendering resources (Lee et al., 2024; Papantonakis et al., 2024; Kerbl et al., 2024; Shi et al., 2025; Windisch et al., 2026). These developments motivate an alternative approach to game-asset transfer: rather than reproducing every source asset and renderer-specific dependency separately, we seek to distill the visual appearance produced by the source engine into a unified Gaussian representation. We refer to this process as *Gaussianizing* a game world.

In order to Gaussianize a game world as a reusable asset supporting six-degree-of-freedom (6DoF) exploration within a valid camera domain, two requirements must be satisfied. First, *spatial observability*: under a limited acquisition budget, relevant surfaces must be observed at sufficient projected resolution and from complementary directions. Second, *radiometric consistency*: training views should encode scene brightness under a shared exposure convention rather than introduce conflicting supervision through per-view exposure changes. Existing large-scale 3DGS methods have substantially improved scalability through spatial partitioning, parallel optimization, and hierarchical

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

representations (Liu et al., 2024; Jiang et al., 2025; Kerbl et al., 2024; Windisch et al., 2026; Wang et al., 2026; Li et al., 2026), but generally do not consider how to faithfully reproduce the source game world's appearance.

Regarding spatial observability, existing large-scale reconstruction pipelines and benchmarks are typically built around prescribed trajectories (Li et al., 2023; Jiang et al., 2025). Street-level views, for example, are often manually captured along roads at approximately fixed camera heights and restricted viewing directions, providing limited observations of surfaces outside the street-level viewing manifold. Moreover, evaluation views are commonly sampled from or near the same acquisition manifold, leaving poorly observed surfaces largely untested. A model may consequently perform well under trajectory-based evaluation while degrading substantially at valid but genuinely off-trajectory viewpoints (Li et al., 2025; Kim et al., 2025). The central problem is therefore not simply to render more training images, but to allocate a fixed rendering budget across camera viewpoints so as to provide sufficient supervision throughout the 6DoF navigation domain.

Recent view-selection methods show that the utility of a camera depends on the observations already available. POp-GS (Wilson et al., 2025) scores each candidate by its expected reduction of parameter uncertainty in the current 3DGS model, and COVER (Chen et al., 2026) favors cameras that observe insufficiently covered surfaces or provide complementary viewing directions. However, they estimate view utility from an initialized or partially optimized representation. Surfaces missing from that intermediate model may not be evaluated reliably. Game-world Gaussianization instead provides access to the source geometry, allowing candidate views to be evaluated directly in terms of surface visibility, projected resolution, and viewing direction.

Access to source geometry also raises a natural question: why not convert the mesh directly into Gaussians? Mesh2Splat (Scolari, 2025) directly converts mesh geometry and a limited set of PBR attributes into Gaussian primitives, but still relies on target-side shading and therefore does not preserve the complete engine-rendered appearance. More closely related, Mesh2GS (Zhu et al., 2026) derives camera and Gaussian sampling densities for global-illumination caching, but assumes uniform light-field sampling and retains mesh-derived albedo buffers and material features rather than producing a self-contained Gaussian asset. These methods address primitive conversion or specialized rendering rather than complete engine-rendered appearance distillation. We therefore use the source mesh to guide acquisition and the source renderer to provide appearance supervision.

Beyond spatial coverage, faithfully reconstructing game environments also requires preserving their radiometric appearance. Modern game scenes exhibit substantial luminance variation across sunlit outdoor areas, deep shadows, and interiors, for which game engines typically employ automatic exposure during navigation. Applying this exposure to training views changes the displayed brightness of the same surface according to camera direction and exposure history, creating conflicting supervision across views. Fixing the LDR exposure avoids such per-view variation but cannot accommodate the full dynamic range. MatrixCity demonstrates this difficulty even without indoor scenes: a single fixed exposure underexposes its outdoor street-level images while overexposing its aerial images, prompting the adoption of automatic exposure for the dataset (Li et al., 2023). Later large-scale Gaussian methods evaluated on MatrixCity, such as Octree-GS (Ren et al., 2025), and Horizon-GS (Jiang et al., 2025), introduce learned appearance embeddings to absorb cross-view exposure and illumination variations. While such modules can mitigate the inconsistencies introduced by auto-exposed LDR inputs to some extent, they do not support the game engine in reproducing the automatic-exposure behavior during runtime navigation.

Based on these observations, we introduce **Game2GS**, a framework that combines coverage-aware view acquisition with HDR appearance distillation. Engine-native probes and their distance maps provide candidate camera poses, while the source mesh estimates which surfaces each candidate observes and quality. We allocate the image budget to views that improve surface coverage in projected resolution, viewing direction, and camera position. At the selected poses, the source renderer engine provides scene-linear HDR images and depth under a shared fixed exposure. We distil these observations into an HDR Gaussian representation using a formulation compatible with different large-scene reconstruction backbones. In this work, we integrate it with both Horizon-GS (Jiang et al., 2025) and Urban-GS (Wang et al., 2026). At runtime, an engine-exported exposure-compensation curve adjusts the displayed brightness during navigation. Together, these components shift large-scale reconstruction from reproducing a set of observed trajectories toward distilling the source game world as a reusable visual asset. Our contributions are summarized as following:

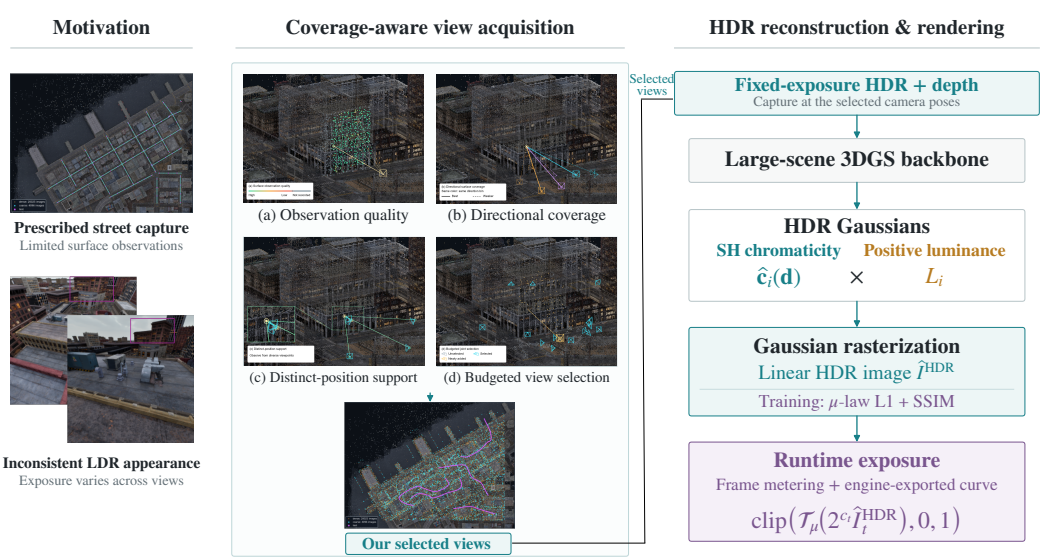

Figure 1: **Overview of Game2GS.** Capture from prescribed street positions at fixed heights produces redundant observations and coverage gaps, while auto-exposed LDR images introduce inconsistent brightness across views. **Coverage-aware view acquisition:** We assess how well each candidate view observes scene surfaces (a). Then select views that add useful observations from different viewing directions (b) and camera positions (c), subject to a fixed image budget (d). **HDR reconstruction and rendering:** Fixed-exposure HDR images and depth supervise HDR Gaussian reconstruction using a large-scene 3DGS backbone. An engine-exported curve controls runtime exposure compensation before display mapping.

- We develop a geometry-guided acquisition strategy that selects complementary, high-quality surface observations under a fixed image budget, before Gaussian reconstruction.
- We acquire scene-linear HDR observations under a fixed exposure and reconstruct an HDR Gaussian representation, deferring exposure adaptation and tone mapping to rendering.
- We establish a free-view evaluation protocol that complements conventional trajectory interpolation with unrestricted valid 6DoF views rendered directly from the source engine.

2 Related Work

Large-Scale Scene Reconstruction. Large-scale neural scene reconstruction addresses memory and optimization costs through spatial decomposition, distributed training, and hierarchical rendering. VastGaussian (Lin et al., 2024) and CityGaussian (Liu et al., 2024) optimize spatial partitions independently, RetinaGS (Li et al., 2024) and CityGS-X (Gao et al., 2025) distribute training across multiple GPUs, and hierarchical or out-of-core methods organize and stream Gaussians according to viewing scale (Kerbl et al., 2024; Windisch et al., 2026). Horizon-GS further combines aerial and street-level imagery through coarse-to-fine training and camera-distribution balancing (Jiang et al., 2025). While these methods substantially improve scalability, they generally treat the input imagery as fixed. Their focus is how to reconstruct a prescribed image collection, rather than whether that collection observes the navigable scene with sufficient visibility, projected resolution, and directional diversity. Consequently, denser street trajectories combined with aerial views do not necessarily provide adequate supervision beyond their camera manifolds.

View Selection for 3D Reconstruction. View planning seeks a compact set of observations that satisfies surface coverage and measurement-quality requirements (Scott et al., 2003). Recent radiance-field methods rank candidate views using predictive uncertainty or information gain, while active mapping systems additionally consider spatial exploration and directional coverage deficits (Pan et al., 2022; Jiang et al., 2024; Wilson et al., 2025; Jin et al., 2025; Chen et al., 2025; 2026). These methods estimate view utility from a partially reconstructed scene representation, tying selection to the current reconstruction state. In our setting, source geometry is available before acquisition. We therefore optimize effective surface coverage directly over the target mesh under a fixed image budget, considering visibility, projected resolution, and viewing-direction diversity without requiring an intermediate radiance field.

HDR Appearance Modeling and Exposure. HDR novel-view synthesis differs in how scene radiance is recovered and how it is mapped to display images. HDR-NeRF recovers radiance from multi-exposure LDR observations by modeling the imaging process (Huang et al., 2022). HDR-GS, Cinematic Gaussians, and GaussHDR extend HDR reconstruction to Gaussian representations with exposure-dependent rendering or tone mapping (Cai et al., 2024; Wang et al., 2024; Liu et al., 2025). LCD-GS factorizes Gaussian appearance into luminance and chromaticity and optimizes against HDR images (Zhang et al., 2025). Our setting provides linear HDR observations of large game worlds under a common fixed exposure. We adapt this appearance factorization to anchor-based large-scene reconstruction and retain a consistent radiance scale across views. Engine-derived exposure compensation is applied to the display image at runtime, after HDR Gaussian rendering.

Mesh-to-Gaussian Scene Conversion. When source geometry is available, it can directly support Gaussian construction. Mesh2Splat converts triangle geometry and selected PBR attributes into surface-aligned Gaussians without image-based optimization, but still relies on target-side shading rather than preserving the complete engine-rendered appearance (Scolari, 2025). Mesh2GS uses known geometry to guide camera and Gaussian sampling for global-illumination caching, but retains mesh-derived albedo and material features (Zhu et al., 2026). In contrast, we use the source mesh only to guide acquisition and distill the complete engine-rendered appearance into a self-contained HDR Gaussian representation, without retaining the original material and lighting pipeline.

3 PRELIMINARIES

3D Gaussian Splatting. 3DGS represents a scene as a set of anisotropic Gaussians $\mathcal{G} = \{G_i\}_{i=1}^{N}$. Each Gaussian has a center μ_i, covariance $\Sigma_i = R_i S_i S_i^\top R_i^\top$, opacity o_i, and view-dependent color c_i. Its spatial kernel is

$$G_i(X) = \exp\left(-\frac{1}{2}(X-\mu_i)^\top \Sigma_i^{-1}(X-\mu_i)\right). \tag{1}$$

After projection, Gaussians are depth-sorted and alpha-composited to render pixel x:

$$C(x) = \sum_i c_i\, \alpha_i(x) \prod_{j<i}\big(1-\alpha_j(x)\big), \tag{2}$$

where $\alpha_i(x) = o_i G_i^{2D}(x)$ and G_i^{2D} is the projected Gaussian.

Engine-native Spatial Probes. Spatial lighting probes are discrete samples distributed throughout a scene to cache illumination, allowing lighting at nearby points to be approximated through interpolation. Cascaded lighting volumes (CLV) organize these samples at multiple spatial scales to support illumination queries over large environments (Majercik et al., 2021). In geometry-guided placement, probes are allocated on spatial grids around scene surfaces, concentrating samples near geometry while retaining a range of spatial resolutions. This organization provides a structured set of spatial sampling locations whose distribution reflects the scene layout.

Probes can also carry directional distance maps that summarize surrounding geometry through an octahedral parameterization, which maps directions on the unit sphere to a compact planar array (Majercik et al., 2019). These maps complement probe positions with approximate distances to surrounding surfaces in different directions, providing geometric cues for evaluating candidate viewpoints. When available from an existing lighting bake, probe positions and their associated distance maps can be reused without repeating probe placement or distance-map computation.

4 GAME2GS

4.1 PROBLEM FORMULATION

We consider a source game world whose mesh geometry, materials, lighting configuration, and renderer are available. Let $\Omega \subseteq \mathrm{SE}(3)$ denote the valid camera-pose domain. Given an image budget B, we select training poses $\mathcal{V} \subseteq \Omega$, $|\mathcal{V}| = B$, and render scene-linear HDR images to reconstruct a Gaussian representation for free-view rendering within Ω and runtime exposure adaptation. Achieving this goal requires deciding both where to capture the training views and how to preserve scene brightness in the captured images.

Sampling manually designed trajectories at a few fixed camera heights can produce many redundant observations while leaving other parts of the scene poorly observed. Increasing model capacity or training time cannot fully compensate for these gaps. We therefore use the known mesh to plan views that observe relevant surfaces from complementary directions and distinct camera positions (*spatial observability*).

Game engines typically adjust exposure during navigation to accommodate large changes in scene brightness. Capturing training images with this auto-exposure introduces brightness changes across views that provide inconsistent supervision. Using a fixed exposure for LDR images avoids these changes, but leaves dark regions underexposed or bright regions saturated. We therefore capture scene-linear HDR images at a fixed exposure, reconstruct an HDR Gaussian representation, and apply auto-exposure and tone mapping at runtime (*radiometric consistency*).

4.2 Geometry-Guided View Acquisition

Under a fixed image budget, a candidate view is valuable when it provides useful observations beyond those already selected (Wilson et al., 2025; Chen et al., 2026). Existing next-best-view methods estimate this value using an intermediate reconstruction. In our setting, the source mesh is available before acquisition. We therefore use it to evaluate candidate views and select the complete training set before Gaussian reconstruction begins.

For each candidate, we use the source mesh to estimate which surfaces it can observe and the quality of those observations. Visibility alone is insufficient: a visible surface may occupy few pixels, be viewed at an oblique angle, or already be well observed from similar poses. Our acquisition objective therefore rewards high-quality observations from complementary viewing directions and qualified support from distinct camera positions. These geometric estimates guide where to capture images, while the source renderer supplies their appearance.

Probe-Based Camera Proposals. A dense search over the six-dimensional camera-pose space is impractical for game worlds. We reuse the global-illumination probes generated by MagicDawn as a source of candidate camera centres. These probes span different locations and heights throughout the three-dimensional scene volume. Their associated DistanceMaps provide approximate distances to surrounding geometry in different directions, which we use to reject positions with insufficient clearance and propose orientations towards nearby surfaces. We then validate the proposed poses using engine-side geometric and scene-content checks, and use source-mesh occlusion tests to determine which surfaces each candidate could observe. The validated poses form a finite candidate set $\mathcal{C} \subseteq \Omega$, from which we select B training views using the coverage objective described below.

Surface Observation Quality. We discretise the target mesh into area-weighted surface elements $\mathcal{P} = \{p_i\}_{i=1}^{M}$, where each element has position $\mathbf{x}_i$, normal $\mathbf{n}_i$, and representative area a_i. For a candidate camera c_j, we evaluate whether its image would provide a useful observation of p_i, rather than counting every visible surface equally. We define the observation quality as

$$q_{ij} = v_{ij}\ \mathrm{clip}\left(w_{ij}^{\mathrm{range}}\,\alpha_{ij}\,\rho_{ij}^{1/2}, 0, 1\right), \tag{3}$$

where $v_{ij} \in \{0,1\}$ denotes visibility after source-mesh occlusion testing, w_{ij}^{range} is the range-band weight, $\alpha_{ij} \in [0,1]$ measures incidence-angle quality, and $\rho_{ij} \in [0,1]$ measures projected resolution. The resolution term is normalised by a target projected area and saturates at 1. Thus, a surface receives little credit when it is occluded or poorly resolved, while increasing its image footprint beyond the target does not provide unbounded reward.

Directional Surface Coverage. A high-quality observation from one direction does not necessarily constrain appearance at substantially different viewpoints. We therefore track how each surface is observed across viewing directions. For each surface element p_i, we express the surface-to-camera direction in a local frame aligned with $\mathbf{n}_i$ and quantise it into direction bins. Given a selected camera set $\mathcal{V}$, we retain the highest-quality observation in each bin:

$$Q_{ib}(\mathcal{V}) = \max_{\substack{c_j \in \mathcal{V} \\ b_i(j)=b}} q_{ij}, \qquad F_{\mathrm{dir}}(\mathcal{V}) = \frac{1}{\sum_i a_i} \sum_i a_i \min\left(1, \frac{1}{L}\sum_b Q_{ib}(\mathcal{V})\right), \tag{4}$$

where $b_i(j)$ is the viewing-direction bin of c_j relative to p_i, and $Q_{ib}(\mathcal{V}) = 0$ if the selected cameras provide no observation in bin b. The parameter L specifies the accumulated directional quality at

which a surface's contribution saturates. Within a bin, a new image contributes only if it improves the best observation. Observations from different bins accumulate until saturation.

Distinct-Position Support. Directional coverage alone does not ensure that a surface is observed from multiple camera positions. Different orientations at the same camera centre may produce additional images, but they do not provide a new positional observation of that surface. Let $\mathbf{t}_j$ denote the centre of camera c_j. For each surface element indexed by i, we collect the distinct centres that observe it with quality above a threshold τ:

$$\mathcal{U}_i(\mathcal{V}) = \{\mathbf{t}_j \mid c_j \in \mathcal{V},\ q_{ij} \geq \tau\}, \qquad F_{\text{sup}}(\mathcal{V}) = \frac{1}{\sum_i a_i} \sum_i a_i \min\left(1, \frac{|\mathcal{U}_i(\mathcal{V})|}{R}\right). \tag{5}$$

Multiple orientations at the same centre thus count only once for a given surface, while observations from additional centres contribute until the target support count R is reached.

Budgeted Joint Selection. The two coverage terms address complementary acquisition needs: F_{dir} rewards useful observations across viewing directions, while F_{sup} prevents those observations from being concentrated at too few camera positions. We combine them with weight λ for distinct-position support:

$$J(\mathcal{V}) = F_{\text{dir}}(\mathcal{V}) + \lambda F_{\text{sup}}(\mathcal{V}), \tag{6}$$

Starting from $\mathcal{V}_0 = \varnothing$, at each step t we select the candidate with the largest spatially adjusted marginal gain:

$$c_t^* = \arg\max_{c \in \mathcal{A}_t} s(c \mid \mathcal{V}_t)\left[J(\mathcal{V}_t \cup \{c\}) - J(\mathcal{V}_t)\right], \tag{7}$$

and update $\mathcal{V}_{t+1} = \mathcal{V}_t \cup \{c_t^*\}$. Here, $\mathcal{A}_t \subseteq \mathcal{C} \setminus \mathcal{V}_t$ contains candidates with positive joint marginal gain that pass the coverage-similarity filter and satisfy the applicable camera-height quota. The factor $s(c \mid \mathcal{V}_t) \in (0, 1]$ softly reduces the scores of candidates in spatially concentrated regions. Selection continues until $|\mathcal{V}| = B$.

A candidate can therefore improve the selected set by revealing a previously uncovered surface, improving the quality of an existing observation, adding a complementary viewing direction, or providing qualified support from another camera position. The height quotas ensure that the selected set also represents the elevated portions of Ω. Because feasibility and the spatial factor depend on the current selection, this is a state-dependent greedy policy. Parameter settings and acceleration details are provided in the Supplementary Material.

4.3 HDR Reconstruction and Runtime Exposure

The selected views provide linear HDR observations under a fixed exposure, preserving radiance differences between bright and dark regions of the game world. The model must fit the large range of HDR values across sunlit and shaded areas while capturing how Gaussian colors vary with viewing direction. To this end, we add a luminance branch to the anchor attribute decoder that predicts a positive scalar for each offset Gaussian. Multiplying this scalar by normalized, view-dependent SH chromaticity (Zhang et al., 2025) gives each Gaussian a linear HDR color. These colors are then alpha-composited to produce the rendered HDR image.

Anchor-Decoded Luminance. Let $\mathbf{f}_a$ denote the feature of anchor a, which generates K offset Gaussians. A two-layer MLP outputs one luminance parameter for each offset:

$$\boldsymbol{\ell}_a = \phi_{\text{lum}}(\mathbf{f}_a) \in \mathbb{R}^K, \qquad L_{a,k} = \exp([\boldsymbol{\ell}_a]_k), \quad k = 1, \ldots, K. \tag{8}$$

Here, $[\boldsymbol{\ell}_a]_k$ is the MLP output for the k-th offset Gaussian, and the exponential ensures that its luminance $L_{a,k}$ is positive.

HDR Photometric Supervision. To reduce the disproportionate influence of high-radiance pixels, we apply the same μ-law transform to the rendered and reference HDR images and compute the L1 and SSIM photometric losses in the compressed domain (Zhang et al., 2025). Depth supervision and other regularisers follow the chosen reconstruction backbone.

Engine-Derived Runtime Exposure. At runtime, we compute the geometric mean luminance $\bar{L}_t$ over valid pixels of each rendered HDR frame. A metering conversion $\mathcal{M}$ maps $\bar{L}_t$ to the EV100 coordinate x_t of the engine-exported exposure-compensation curve $\mathcal{C}$, which returns a compensation

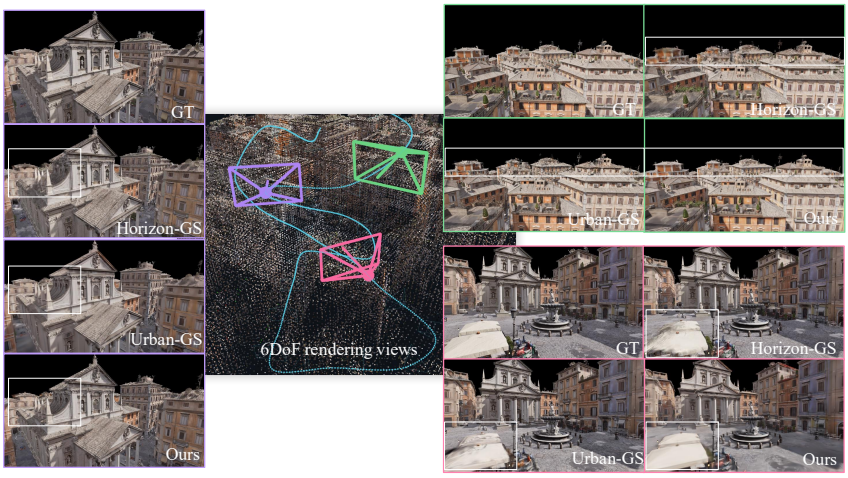

Figure 2: Qualitative comparison on independent *free-6DoF* test views of *Rome*. Colored camera frusta indicate the viewpoints corresponding to the matching image borders. Horizon-GS and Urban-GS are trained on LDR images acquired using the *prescribed-view* strategy, while Game2GS (Ours) uses Urban-GS with fixed-exposure HDR images acquired using our *coverage-aware* strategy. Compared with the baselines, our approach better preserves rooflines, facade details, and foreground structures, with fewer rendering artifacts in the highlighted regions. GT denotes reference images rendered by the source engine.

c_t in stops: $x_t = \mathcal{M}(\bar{L}_t), c_t = \mathcal{C}(x_t)$. For display, we scale the linear HDR render by 2^{c_t} before applying our μ-law mapping:

$$\hat{\mathbf{I}}_t^{\text{disp}} = \text{clip}\Big(\mathcal{T}_\mu(2^{c_t}\hat{\mathbf{I}}_t^{\text{HDR}}), 0, 1\Big)\,. \tag{9}$$

5 Experiments

Datasets. We evaluate our method on four large-scale scenes divided into two groups. The first comprises two existing benchmark scenes: *Block A* from MatrixCity (Li et al., 2023) and *Elvenruins* from Horizon-GS (Jiang et al., 2025). These are the two benchmark scenes for which we could obtain the source Unreal Engine assets needed for view acquisition. The second comprises two private game-world scenes, *Rome* and *Arena*, featuring complex geometry, severe occlusions, and substantial luminance variations across outdoor areas, interiors, shadows, and enclosed spaces. MatrixCity provides sparse and dense street-view training sets, with the latter sampling the same prescribed routes more densely (Li et al., 2023). For *Block A*, we match our street-image counts to these two sets, corresponding to B_{low} and B_{high}, respectively, while keeping the aerial images fixed.

Test Views. We evaluate novel-view rendering on two complementary test sets. The *trajectory-near* set follows the evaluation protocols of MatrixCity (Li et al., 2023) and Horizon-GS (Jiang et al., 2025), using held-out cameras along or near the prescribed trajectories. The *free-6DoF* set comprises independently sampled cameras throughout the valid navigation domain, covering positions, heights, and viewing directions beyond those of the training views.

Evaluation Metrics. We report PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018) to assess novel-view rendering quality. For HDR evaluation, we apply the same fixed μ-law display transform to both rendered and ground-truth images before computing these metrics.

Implementation Details. We integrate Game2GS with two large-scale reconstruction backbones, Horizon-GS (Jiang et al., 2025) and Urban-GS (Wang et al., 2026), retaining their respective core reconstruction pipelines. Both implementations incorporate the HDR color parameterization and photometric supervision described in Sec. 4.3. HDR models are trained on fixed-exposure images, with photometric losses computed after μ-law compression. Exposure compensation is used only for runtime display; quantitative HDR evaluation uses the fixed display transform specified above.

The *prescribed-view* baseline uses cameras along predefined street routes, while our *coverage-aware* strategy selects camera poses using the geometry-guided objective in Sec. 4.2. When comparing acquisition strategies or HDR variants within the same backbone, we keep the remaining reconstruc-

Table 1: Quantitative comparison on independent *free-6DoF* test views. Hier-GS, Horizon-GS, and Urban-GS are trained on LDR images captured using the *prescribed-view* strategy. Game2GS uses Urban-GS as its reconstruction backbone and is trained on fixed-exposure HDR images acquired using our *coverage-aware* strategy. Each entry reports PSNR↑ / SSIM↑ / LPIPS↓.

Method	Elvenruins	MatrixCity	Rome	Arena
Hier-GS	18.463 / 0.5370 / 0.4772	15.690 / 0.5450 / 0.5518	20.877 / 0.6458 / 0.3941	19.662 / 0.5912 / 0.5355
Horizon-GS	18.788 / 0.5379 / 0.4671	18.661 / 0.6404 / 0.4234	22.295 / 0.6691 / 0.4129	21.602 / 0.5946 / 0.5381
Urban-GS	19.159 / 0.5438 / 0.4517	20.019 / 0.6622 / 0.3959	23.500 / 0.6900 / 0.3850	23.443 / 0.6663 / 0.4522
Game2GS (Ours)	**23.284 / 0.6728 / 0.3741**	**25.767 / 0.8176 / 0.3315**	**26.968 / 0.7419 / 0.3761**	**28.368 / 0.8312 / 0.3354**

Table 2: Controlled comparison of street-level acquisition strategies on part region of MatrixCity *Block A* using Urban-GS with our HDR formulation. At each image budget, both strategies use identical aerial images, fixed-exposure HDR capture settings, and training schedules, differing only in street-view selection. Evaluation uses shared *trajectory-near* and *free-6DoF* test views.Each entry reports PSNR↑ / SSIM↑ / LPIPS↓.

Street-level acquisition	Trajectory-near B_{low}	Trajectory-near B_{high}	Free-6DoF B_{low}	Free-6DoF B_{high}
Prescribed views	**26.866 / 0.8823 / 0.2909**	**28.053 / 0.8963 / 0.2745**	20.554 / 0.7341 / 0.4066	20.105 / 0.7063 / 0.4395
Ours coverage-aware views	25.668 / 0.8278 / 0.3634	27.615 / 0.8794 / 0.2981	**25.767 / 0.8176 / 0.3315**	**26.182 / 0.8294 / 0.3235**

tion settings fixed and provide identical ground-truth depth supervision to all applicable methods. Backbone-specific training schedules are detailed in the supplementary material.

5.1 Main Reconstruction Comparison

We evaluate reconstruction quality for *6DoF free-view exploration* beyond prescribed capture trajectories, using independent test views with varied camera positions and orientations within the valid camera domain. Table 1 compares Hier-GS (Kerbl et al., 2024), Horizon-GS (Jiang et al., 2025), Urban-GS (Wang et al., 2026), and Game2GS. The baselines use prescribed-view LDR training, while Game2GS combines coverage-aware acquisition and fixed-exposure HDR reconstruction using Urban-GS as its backbone. Under this free-6DoF evaluation, Game2GS improves PSNR over Urban-GS by 4.13, 5.75, 3.47, and 4.93 dB on Elvenruins, MatrixCity, Rome, and Arena, respectively. It also achieves higher SSIM and lower LPIPS. Figure 2 further shows clearer architectural details and fewer artifacts around rooflines and foreground structures across different viewing heights and directions, demonstrating improved rendering fidelity for 6DoF exploration.

5.2 Controlled Acquisition Comparison

Table 2 isolates the effect of the acquisition strategy under matched image budgets. As expected, prescribed-view acquisition performs better on trajectory-near test views, since these views largely interpolate the same camera trajectories used for training, but its performance degrades substantially on independent free-6DoF views. Moreover, increasing the density of prescribed views along the same trajectories brings limited improvement on free-6DoF evaluation, indicating that denser sampling of an already observed camera manifold does not substantially improve coverage of the broader navigable scene. In contrast, our coverage-aware acquisition sacrifices some accuracy on trajectory-near views but achieves better free-6DoF reconstruction; importantly, increasing its acquisition budget further improves both trajectory-near and free-6DoF performance, with a larger gain on free-6DoF views, suggesting that additional views are allocated to complementary observations that improve reconstruction across the scene.

5.3 Radiometric Representation Ablation

Radiometric ablation. We evaluate radiometric design choices on our *Arena* game scene, where bright, sunlit exteriors connect to dim interiors, creating strong brightness differences and pronounced shadow boundaries. We conduct this ablation with Horizon-GS to assess applicability beyond the Urban-GS backbone used in our main experiments. Table 3 shows that our fixed-exposure

Table 3: Radiometric ablation on the *Arena* scene. Our fixed-exposure HDR approach achieves the best overall results, benefiting from consistent radiometric supervision and preserved scene dynamic range. Adding an appearance module to auto-exposed LDR yields worse results when evaluated using the mean training embedding without test-view optimization. We conduct this ablation with Horizon-GS to assess applicability beyond the Urban-GS backbone used in our main experiments. Each entry reports PSNR↑ / SSIM↑ / LPIPS↓.

Capture and representation	App_dim = 32	Farmland
Auto-exposed LDR		23.646 / 0.7543 / 0.4179
Auto-exposed LDR	✓	22.580 / 0.7480 / 0.4289
Fixed-exposure LDR		25.981 / 0.8125 / 0.3811
Fixed-exposure HDR		**27.790 / 0.8137 / 0.3536**



Figure 3: Qualitative radiometric comparison on *Arena* using Horizon-GS. From left to right: auto-exposed reference, fixed-exposure LDR, auto-exposed LDR without and with a 32-dimensional appearance embedding, and our HDR rendering using the recorded reference exposure. Our approach better reproduces the reference appearance: the interior remains dark when viewed from the bright exterior (top row), while increased exposure reveals indoor details after the camera enters the room (bottom two rows).

HDR approach achieves the best overall results, improving PSNR by 4.144 dB over auto-exposed LDR and by 1.809 dB over fixed-exposure LDR, although the latter yields slightly higher SSIM.

Figure 1 illustrates the desired exposure behavior and the differences between reconstruction settings. When looking into the room from the bright exterior (top row), the reference exposure accommodates the outdoor brightness, leaving the interior relatively dark. After the camera enters the room (bottom two rows), exposure increases, revealing the interior walls, furniture, and shadowed details. This change reflects display exposure adaptation under unchanged scene lighting.

Fixed-exposure LDR loses details in bright or dark regions that cannot be recovered during rendering, while auto-exposed LDR introduces inconsistent brightness supervision across views and artifacts near indoor–outdoor boundaries. Adding an appearance embedding does not fully resolve this issue, as the mean training embedding does not provide viewpoint-appropriate exposure at test time. In contrast, our fixed-exposure HDR approach preserves scene radiance on a shared scale and applies exposure separately during rendering. Using the recorded reference exposure without test-time optimization, it better reproduces both viewing conditions: interiors remain dark when viewed from outside and become visible under the higher exposure used inside.

6 Conclusion

We presented **Game2GS**, a framework for Gaussianizing game worlds into reusable HDR representations for 6DoF free-view rendering. Game2GS combines geometry-guided, coverage-aware view acquisition under a fixed image budget with fixed-exposure HDR reconstruction, enabling complementary surface observations while preserving scene radiance on a consistent scale. Experiments on benchmark and game-world scenes demonstrate improved reconstruction quality at independent 6DoF viewpoints and more faithful appearance under varying runtime exposure. Our results suggest a path toward distilling complex game worlds into reusable Gaussian assets, future work extending the framework to dynamic content, relighting, and efficient deployment.

AI USE STATEMENT

In this work, we used generative AI tools to assist in developing and refining conceptual frameworks and hypotheses, formulating mathematical claims and derivations, designing research methodology and experiments, implementing methods, and interpreting experimental results. We did not use generative AI tools to generate synthetic datasets, translate research materials, clean or reformat datasets, or conduct qualitative or thematic data analysis. Tasks related to formal mathematical proofs were not applicable to this work. Additionally, we used generative AI tools for code assistance, literature search and analysis, brainstorming, drafting, and editing the manuscript. All AI-assisted work was reviewed by the authors, with code tested for correctness and written content checked against the relevant sources and experimental results. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Tomas Akenine-Möller, Eric Haines, Naty Hoffman, Angelo Pesce, Michał Iwanicki, and Sébastien Hillaire. *Real-Time Rendering 4th Edition*. A K Peters/CRC Press, Boca Raton, FL, USA, 2018. ISBN 978-1-13862-700-0.

Yuanhao Cai, Zihao Xiao, Yixun Liang, Minghan Qin, Yulun Zhang, Xiaokang Yang, Yaoyao Liu, and Alan Yuille. Hdr-gs: Efficient high dynamic range novel view synthesis at 1000x speed via gaussian splatting. In *Advances in Neural Information Processing Systems*, volume 37, pp. 68453–68471, 2024.

Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu. Activegamer: Active gaussian mapping through efficient rendering. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16486–16497. IEEE, 2025.

Timothy Chen, Adam Dai, Maximilian Adang, Grace Gao, and Mac Schwager. Coverage optimization for camera view selection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19443–19451, 2026.

Yuanyuan Gao, Hao Li, Jiaqi Chen, Zhengyu Zou, Zhihang Zhong, Dingwen Zhang, Xiao Sun, and Junwei Han. Citygs-x: A scalable architecture for efficient and geometrically accurate large-scale scene reconstruction. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27187–27196. IEEE, 2025.

Jason Gregory. *Game Engine Architecture*. A K Peters/CRC Press, 3 edition, 2018. ISBN 978-1-138-03545-4.

Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18377–18387. IEEE, 2022.

Lihan Jiang, Kerui Ren, Mulin Yu, Linning Xu, Junting Dong, Tao Lu, Feng Zhao, Dahua Lin, and Bo Dai. Horizon-gs: Unified 3d gaussian splatting for large-scale aerial-to-ground scenes. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26789–26799. IEEE, 2025.

Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and mapping with radiance fields using fisher information. In *European conference on computer vision*, pp. 422–440. Springer, 2024.

Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. Activegs: Active scene reconstruction using gaussian splatting. *IEEE Robotics and Automation Letters*, 10 (5):4866–4873, 2025.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4), 2023.

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.

Minsu Kim, Subin Jeon, In Cho, Mijin Yoo, and Seon Joo Kim. Exploregs: Explorable 3d scene reconstruction with virtual camera samplings and diffusion priors. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27042–27051. IEEE, 2025.

Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3d gaussian representation for radiance field. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21719–21728. IEEE, 2024.

Bingling Li, Shengyi Chen, Luchao Wang, Kaimin Liao, Sijie Yan, and Yuanjun Xiong. Retinags: Scalable training for dense scene rendering with billion-scale 3d gaussians. *arXiv preprint arXiv:2406.11836*, 2024.

Changbai Li, Haodong Zhu, Hanlin Chen, Xiuping Liang, Tongfei Chen, Shuwei Shao, Linlin Yang, Huobin Tan, and Baochang Zhang. Urbangs: Efficient and scalable architecture for geometrically accurate large-scene reconstruction. In *International Conference on Learning Representations*, volume 2026, pp. 9733–9758, 2026.

Hao Li, Chenming Wu, Ming Yuan, Yan Zhang, Chen Zhao, Chunyu Song, Haocheng Feng, Errui Ding, Dingwen Zhang, and Jingdong Wang. Xld: A cross-lane dataset for benchmarking novel driving view synthesis. In *2025 International Conference on 3D Vision (3DV)*, pp. 659–669. IEEE, 2025.

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 *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3182–3192. IEEE, 2023.

Yuzhi Liang, Qi Song, Rui Wang, Yuchi Huo, and Hujun Bao. Automatic mesh and shader level of detail. *IEEE Transactions on Visualization and Computer Graphics*, 29(10):4284–4295, 2023. doi: 10.1109/TVCG.2022.3188775.

Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, and Wenming Yang. Vastgaussian: Vast 3d gaussians for large scene reconstruction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5166–5175. IEEE, 2024.

Jinfeng Liu, Lingtong Kong, Bo Li, and Dan Xu. Gausshdr: High dynamic range gaussian splatting via learning unified 3d and 2d local tone mapping. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5991–6000. IEEE, 2025.

Yang Liu, Chuanchen Luo, Lue Fan, Naiyan Wang, Junran Peng, and Zhaoxiang Zhang. Citygaussian: Real-time high-quality large-scale scene rendering with gaussians. In *European Conference on Computer Vision*, pp. 265–282. Springer, 2024.

Zander Majercik, Jean-Philippe Guertin, Derek Nowrouzezahrai, and Morgan McGuire. Dynamic diffuse global illumination with ray-traced irradiance fields. *Journal of Computer Graphics Techniques*, 8(2):1–30, 2019. URL `https://jcgt.org/published/0008/02/01/`.

Zander Majercik, Adam Marrs, Josef Spjut, and Morgan McGuire. Scaling probe-based real-time dynamic global illumination for production. *Journal of Computer Graphics Techniques*, 10(2): 1–29, 2021. URL `https://jcgt.org/published/0010/02/01/`.

Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. Activenerf: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision*, pp. 230–246. Springer, 2022.

Panagiotis Papantonakis, Georgios Kopanas, Bernhard Kerbl, Alexandre Lanvin, and George Drettakis. Reducing the memory footprint of 3d gaussian splatting. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–17, 2024.

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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–15, 2025.

Stefano Scolari. Mesh2splat: Fast mesh to 3d gaussian splat conversion. https://github.com/electronicarts/mesh2splat, 2025. Extended and updated version of the author's Master's thesis at KTH Royal Institute of Technology.

William R Scott, Gerhard Roth, and Jean-François Rivest. View planning for automated three-dimensional object reconstruction and inspection. *ACM Computing Surveys (CSUR)*, 35(1):64–96, 2003.

Yuang Shi, Géraldine Morin, Simone Gasparini, and Wei Tsang Ooi. Lapisgs: Layered progressive 3d gaussian splatting for adaptive streaming. In *2025 International Conference on 3D Vision (3DV)*, pp. 991–1000. IEEE, 2025.

Chao Wang, Krzysztof Wolski, Bernhard Kerbl, Ana Serrano, Mojtaba Bemana, Hans-Peter Seidel, Karol Myszkowski, and Thomas Leimkühler. Cinematic gaussians: Real-time hdr radiance fields with depth of field. *Computer Graphics Forum*, 43(7):e15214, 2024.

Meng Wang, Changqun Xia, Yuze Wang, Junyi Wang, Wantong Duan, Xinxiong Xie, and Yue Qi. Urban-gs: A unified 3d gaussian splatting framework for compact and high-fidelity aerial-to-street reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 33207–33216, 2026.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Joey Wilson, Marcelino Almeida, Sachit Mahajan, Martin Labrie, Maani Ghaffari, Omid Ghasemalizadeh, Min Sun, Cheng-Hao Kuo, and Arnab Sen. Pop-gs: Next best view in 3d-gaussian splatting with p-optimality. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3646–3655. IEEE, 2025.

Felix Windisch, Thomas Köhler, Lukas Radl, Mattia D'Urso, Michael Steiner, Dieter Schmalstieg, and Markus Steinberger. A lod of gaussians: Out-of-core training and rendering for seamless ultra-large scene reconstruction. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, SIGGRAPH Conference Papers '26, pp. 1–11. Association for Computing Machinery, 2026.

Kaixuan Zhang, Minxian Li, Mingwu Ren, Jiankang Deng, and Xiatian Zhu. High dynamic range 3d gaussian splatting via luminance-chromaticity decomposition. *arXiv preprint arXiv:2511.12895*, 2025.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

Haoran Zhu, Youcheng Cai, Huangsheng Du, Jingyang Meng, and Ligang Liu. Mesh2gs: White-box 3dgs construction via plenoptic sampling. *arXiv preprint arXiv:2606.21898*, 2026.

A Appendix

A.1 Backbone Integration and HDR Supervision

Backbone integration. We integrate Game2GS into Horizon-GS and Urban-GS while preserving their backbone-specific reconstruction pipelines. Horizon-GS retains its coarse-to-fine optimization, whereas Urban-GS retains global training and local refinement. Our integration introduces the HDR color representation and photometric supervision described in Sec. 4.3, alongside the geometric and regularization terms of each backbone.

For acquisition comparisons within the same backbone, we keep the reconstruction configuration fixed, including the model architecture, optimization schedule, image-processing convention, and depth supervision. The selected views and the geometry initialized from them follow the corresponding acquisition strategy.

HDR representation and input processing. Each Gaussian represents radiance as the product of view-dependent chromaticity and a positive scalar intensity. Chromaticity is obtained from nonnegative, normalized spherical-harmonic outputs, while an additional MLP predicts the log-intensity. A neutral chromaticity is used when the normalization denominator is close to zero. The renderer composites the resulting colors in linear radiance space.

HDR inputs are fixed-exposure, scene-linear images stored in floating-point format. Their radiometric scale is preserved without per-image normalization or learned exposure compensation. Nonfinite inputs are rejected, and negative radiance values are handled by the configured input-validation policy.

Photometric and depth supervision. We apply the same fixed μ-law compression to predicted and reference radiance before computing the photometric loss. The objective combines a masked L_1 term with masked SSIM. The L_1 term is normalized by the number of valid pixels and channels. For SSIM, local statistics are normalized by the valid support within each window, preventing excluded pixels from contributing as zero-valued observations. The compressed radiance is not upper-clipped during training, so over-scale predictions retain a photometric gradient.

All applicable methods receive the ground-truth depth associated with their input views. Depth units and the optical-axis depth convention are accounted for during coordinate conversion, and invalid depth values are excluded. The depth-loss formulation remains backbone-specific; the Urban-GS HDR implementation compares inverse depths over valid support.

A.2 Candidate Generation and Geometric Validation

We use the global-illumination probes generated by MagicDawn to propose camera centers at multiple scales throughout the scene volume. Each probe's DistanceMap approximates distances to nearby geometry along different directions. These estimates are used to reject positions with insufficient predicted clearance and to propose orientations toward scene surfaces.

Because DistanceMap estimates are approximate, candidate poses undergo engine-side geometric and scene-content validation. Source-mesh occlusion queries then determine which surface elements each candidate can observe. The validated poses form the candidate set $\mathcal{C} \subseteq \Omega$.

Probe generation, pose validation, and visibility queries are performed before view selection. The selector operates on the resulting camera–surface observations.

A.3 Surface Representation and Observation Quality

We discretize the target mesh into surface elements $\mathcal{P} = \{p_i\}_{i=1}^{M}$. Each element has a position $\mathbf{x}_i$, normal $\mathbf{n}_i$, and representative geometric area a_i. Area weighting accounts for variation in surface-sampling density. The implementation also supports an optional suitability weight u_i, giving an effective objective weight $\widetilde{a}_i = a_i u_i$. Uniform suitability weights recover the area-weighted objective in the main paper. Projected-area estimation always uses the geometric area a_i.

For a candidate camera c_j with center $\mathbf{t}_j$, we define

$$d_{ij} = \|\mathbf{t}_j - \mathbf{x}_i\|, \qquad \mu_{ij} = \mathbf{n}_i^{\top} \frac{\mathbf{t}_j - \mathbf{x}_i}{d_{ij}}. \tag{10}$$

Camera–surface pairs are first filtered by the camera frustum, a valid distance interval, and an incidence threshold. The range term w_{ij}^{range} is a binary gate in the reference implementation.

For retained pairs, the projected pixel area is approximated by

$$A_{ij}^{\mathrm{px}} \approx a_i \, \mu_{ij} \, f_x f_y \frac{d_{ij}}{z_{ij}^3}, \tag{11}$$

where z_{ij} is camera-space depth and f_x, f_y are focal lengths in pixels. Near the optical axis, the distance factor approaches an inverse-square dependence.

We combine incidence and projected-resolution quality as

$$\alpha_{ij} = \mathrm{clip}\left(\frac{\mu_{ij} - \mu_{\min}}{1 - \mu_{\min}}, 0, 1\right), \qquad \rho_{ij} = \min\left(1, \frac{A_{ij}^{\mathrm{px}}}{A_{\mathrm{target}}}\right). \tag{12}$$

The observation quality is

$$q_{ij} = v_{ij}\ \mathrm{clip}\big(w_{ij}^{\mathrm{range}} \alpha_{ij} \sqrt{\rho_{ij}}, 0, 1\big), \tag{13}$$

where v_{ij} denotes source-mesh visibility. Only visibility-verified observations with positive quality contribute to selection.

A.4 Directional Coverage and Position Support

Directional coverage. For each surface element, we express the surface-to-camera direction in a local frame aligned with its normal and discretize it into azimuth and incidence bins. Within each bin, $Q_{ib}(\mathcal{V})$ retains the highest-quality observation among the selected cameras.

An additional observation in the same bin improves coverage only if its quality exceeds the current maximum. Observations in different bins accumulate until the directional target is reached. This target measures accumulated observation quality rather than requiring an exact number of cameras.

Distinct-position support. Directional coverage alone does not distinguish observations captured from different physical camera centers. We therefore maintain, for each surface element, the set of distinct selected centers providing observations with $q_{ij} \geq \tau$.

Multiple orientations at the same center count only once toward the support of a given surface. The same center may nevertheless support multiple surfaces. Camera centers are deduplicated by equality of their stored coordinates, rather than by merging nearby positions within a spatial radius.

The support contribution saturates at a target number of centers. Additional centers may be retained for diagnostics without increasing the objective. This term encourages observations from multiple positions, but does not impose a minimum baseline or guarantee sufficient parallax.

A.5 Greedy Selection and Spatial Distribution

Starting from an empty set, we evaluate each candidate's joint marginal gain:

$$\Delta_J(c \mid \mathcal{V}) = \Delta_{\mathrm{dir}}(c \mid \mathcal{V}) + \lambda \Delta_{\mathrm{sup}}(c \mid \mathcal{V}). \tag{14}$$

An observation from a new qualified camera center can provide positive support gain even when it does not improve directional coverage. Candidates must have positive joint gain to be selected.

To discourage spatial concentration, we track the occupancy of distinct selected centers in height bands and horizontal cells. Candidates receive a multiplicative penalty when their associated regions become overrepresented. The selection score is

$$\mathrm{score}(c \mid \mathcal{V}) = s(c \mid \mathcal{V})\, \Delta_J(c \mid \mathcal{V}), \tag{15}$$

where $s(c \mid \mathcal{V})$ decreases as regional occupancy exceeds its reference level. Selecting another orientation at an existing center does not increase the unique-position occupancy counts.

The selector additionally applies quotas to low-, middle-, and high-altitude layers at configured budget checkpoints. Layer membership is determined relative to local ground, while occupancy bands are indexed using world height. A candidate is eligible only if its layer has remaining quota at the active checkpoint.

Smaller training budgets are prefixes of the same selected-camera ranking, providing nested view sets for acquisition comparisons. If a checkpoint cannot be filled with eligible positive-gain candidates, the selector reports the shortfall rather than silently relaxing the constraints.

A.6 Support-Aware Redundancy Suppression

We evaluate redundancy between candidates and selected views whose camera centers and forward directions are sufficiently close. Similarity is measured using the surface elements observed above the support-quality threshold, rather than shared directional bins.

Let $S(c)$ denote the qualifying surface set of camera c. The area-weighted containment score is

$$D(c, c') = \frac{\sum_{i \in S(c) \cap S(c')} \widetilde{a}_i}{\min\left(\sum_{i \in S(c)} \widetilde{a}_i,\ \sum_{i \in S(c')} \widetilde{a}_i\right)}. \tag{16}$$

Pairs with zero qualifying area are excluded from this comparison.

A candidate with high containment may be suppressed only when it has zero distinct-position support gain. Consequently, a view from a new qualified center remains eligible while it improves support, even if its observed surfaces substantially overlap those of an existing view.

A.7 Sparse Evaluation and Method Scope

A dense camera–surface matrix would require $|\mathcal{C}||\mathcal{P}|$ entries. We instead construct sparse relations using geometric proposal tests followed by source-scene occlusion queries. The resulting camera-row representation stores surface indices, observation qualities, and directional bins. The selector requires these relations to be marked as occlusion verified.

After selecting a camera, only the surface elements in its sparse row require updates to their directional and support states. Candidate scores are maintained in a lazy priority queue and recomputed when revisited.

Only retained relations contribute to the objective. An absent camera–surface pair may have been excluded by geometric tests, left unqueried, or rejected by an occlusion query; absence alone does not establish invisibility. Support counts therefore refer to the retained sparse observations rather than an exhaustive evaluation of all possible pairs.

The coarse and fine acquisition budgets denote prefixes of the same camera ranking. They do not inherently imply separate surface resolutions or a second visibility pass. The selector adds cameras individually and does not explicitly optimize camera groups, pairwise baselines, or a co-visibility graph. Its objectives encourage directional diversity and qualified multi-position observations without guaranteeing reconstruction accuracy.