URBANBP-GS: BUILDING-AWARE PARTITIONING AND PLANARITY-CONSTRAINED UPDATING FOR MORPHOLOGY-ADAPTIVE URBAN NOVEL-VIEW SYNTHESIS

Anonymous authors
Paper under double-blind review

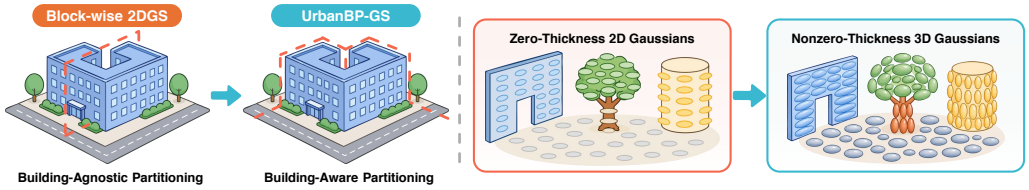

Figure 1: UrbanBP-GS preserves connected-building ownership and retains finite-thickness 3D Gaussians under planarity-constrained updating.

ABSTRACT

Block-wise reconstruction based on 2D Gaussian Splatting (2DGS) has emerged as a promising paradigm for high-fidelity novel-view synthesis of large urban scenes. However, this paradigm encounters bottlenecks in scene partitioning and primitive representation. Partitioning schemes suffer from inter-block primitive imbalance and building fragmentation, while 2DGS limits the rendering of geometrically irregular structures. To address these limitations, we propose UrbanBP-GS, a morphology-adaptive framework that uses building-aware scene partitioning to balance primitive counts and preserve building connectivity, and planarity-constrained 3D Gaussian updating to support high-fidelity urban rendering with irregularity-tolerant primitives. For building-aware scene partitioning, UrbanBP-GS first initializes scene partitioning under a primitive-balance constraint, then detects connected buildings spanning multiple blocks and consolidates each building into its dominant block under a per-block primitive-count variation budget. For planarity-constrained 3D Gaussian updating, UrbanBP-GS associates each planar region with its contributing Gaussians, estimates their normals through soft shortest-axis selection, and promotes orientation consistency among Gaussians in the same planar region while preserving nonzero-thickness primitives to represent irregular structures. Extensive experiments demonstrate that UrbanBP-GS achieves superior PSNR and LPIPS performance across multiple large-scale urban scenes, with particularly notable gains in densely built areas.

1 INTRODUCTION

Large-scale urban novel-view synthesis is widely applied in fields such as digital twins, autonomous-driving simulation, and immersive navigation. 3D Gaussian Splatting (3DGS) provides an effective foundation for this task by representing scenes with explicit anisotropic primitives. Extending 3DGS to city-scale environments remains challenging because it involves optimizing a vast number of Gaussian primitives with a highly nonuniform spatial distribution within a single global model. Large-scale Gaussian reconstruction methods therefore partition the scene into spatial blocks and optimize each block independently (Lin et al., 2024; Liu et al., 2024).

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.

Representative methods use a global coarse representation to guide primitive partitioning, local block optimization, and final model merging (Lin et al., 2024; Liu et al., 2024). Subsequent methods retain block-wise optimization while varying how they form blocks, balance primitive counts, and preserve cross-block consistency (Wu et al., 2025; Hung et al., 2025; Su et al., 2025; Wang et al., 2026). CityGaussianV2 retains the block-wise reconstruction strategy but represents each block with 2D Gaussian Splatting (2DGS), which is well suited to planar urban structures such as facades, roads, and rooftops (Huang et al., 2024; Liu et al., 2025b). Nevertheless, limitations remain in scene partitioning and primitive representation, as illustrated in Figure 1. Existing partitioning schemes are prone to inter-block primitive imbalance and building fragmentation, while the zero-thickness planar primitives used in 2DGS are less flexible when representing geometrically irregular urban structures.

Scene partitioning requires balancing spatial regularity, the distribution of primitives across blocks, and building connectivity. Early coordinate-grid partitioning produces spatially regular blocks but often results in highly imbalanced primitive counts across them. Quantile-cut partitioning balances primitive counts but risks splitting connected buildings and introducing inconsistencies during block fusion. Conversely, preserving building integrity may disrupt the previously established primitive-count balance.

2D Gaussians effectively parameterize regular surfaces (Huang et al., 2024). However, forcing every primitive to be two-dimensional reduces the primitive-level flexibility needed to faithfully render geometrically irregular structures such as vegetation, protruding elements, and cranes. Volumetric 3D Gaussians offer greater freedom for irregular geometry, but they do not inherently provide the surface-oriented inductive bias that benefits 2DGS in planar regions.

In this paper, we propose UrbanBP-GS, a morphology-adaptive framework in which connected buildings guide scene partitioning and planar regions guide per-block 3D Gaussian updating. The building-aware partitioner initializes blocks with comparable primitive counts and refines their ownership using connected-building evidence, reducing fragmentation while keeping block sizes controlled. The planarity-constrained updater groups contributing Gaussians by planar region and promotes consistent surface directions while retaining finite-thickness 3D primitives for irregular urban structures. Experiments on four large-scale urban scenes show that UrbanBP-GS achieves quantitative results competitive with those of state-of-the-art methods. Qualitative comparisons further show that it preserves fine details on planar building surfaces and thin elements in geometrically irregular structures.

The contributions of this work are summarized in the following three points.

- We propose UrbanBP-GS, which uses connected buildings to guide scene partitioning and planar regions to guide per-block 3D Gaussian updating.
- We introduce building-aware scene partitioning, which integrates primitive-count balance with connected-building ownership to reduce building fragmentation across independently optimized blocks.
- We develop planarity-constrained 3D Gaussian updating, which promotes consistent surface directions among Gaussians contributing to the same planar region while retaining a finite-thickness 3D representation for irregular urban structures.

2 Related Work

2.1 Urban Novel-View Synthesis

Urban novel-view synthesis reconstructs large urban scenes from extensive image collections while preserving local appearance and geometry. Early large-scale NeRF methods enabled neural rendering at city scale by decomposing large scenes into smaller regions for independent modeling. Mega-NeRF partitions a scene into independently optimized local radiance fields (Turki et al., 2022). Block-NeRF combines local fields with appearance matching for urban driving sequences (Tancik et al., 2022). Switch-NeRF learns scene-dependent expert assignment (Mi & Xu, 2023). SUDS separates static, dynamic, and far-field components in large urban environments (Turki et al., 2023). Together, these methods have established a practical approach to urban novel-view synthesis. Smaller scene regions are reconstructed independently and then integrated into a complete scene model.

Gaussian-based methods use the same decomposition strategy but replace implicit fields with explicit primitives that support interactive rendering (Kerbl et al., 2023). VastGaussian uses progressive cells and visibility-aware data assignment (Lin et al., 2024), whereas CityGaussian combines a coarse global prior with block refinement and model fusion (Liu et al., 2024). CityGaussianV2 adapts the block-wise reconstruction paradigm to surface-oriented 2DGS (Liu et al., 2025b). Block-Gaussian, LoBE-GS, and HUG explore adaptive blocks, primitive-count balance, and hierarchical aerial reconstruction (Wu et al., 2025; Hung et al., 2025; Su et al., 2025). DOGS coordinates distributed training through shared-Gaussian consensus (Chen & Lee, 2024).

2.2 Scene Partitioning and Surface-Oriented Gaussian Regularization

Block-wise reconstruction partitions a large scene into smaller, spatially compact blocks that can be optimized independently. Fixed geographic or coordinate-based blocks provide clear ownership but do not guarantee comparable primitive counts under nonuniform urban content (Tancik et al., 2022; Liu et al., 2024). Adaptive strategies use learned expert assignment, visibility-aware cells, or visible-Gaussian load objectives (Mi & Xu, 2023; Lin et al., 2024; Hung et al., 2025), whereas boundary methods align local appearance, optimize overlapping regions, or enforce cross-block consensus (Tancik et al., 2022; Chen & Lee, 2024; Wang et al., 2026). Semantic decompositions further show that masks and instances can organize urban structure (Siddiqui et al., 2023; Turki et al., 2023). PG-SAG groups building subregions using segmentation masks and cross-view visibility under a load-balancing objective (Wang et al., 2025). It is designed for surface reconstruction, whereas UrbanBP-GS uses building connectivity to refine primitive ownership subject to a per-block primitive-count variation budget for large-scale urban novel-view synthesis.

Surface-oriented methods represent and regularize scene geometry in different ways. 2DGS represents surfaces with oriented disks (Huang et al., 2024). SuGaR aligns Gaussians with an implicit surface (Guedon & Lepetit, 2024), and Gaussian Opacity Fields derives an opacity field from 3D Gaussians for surface extraction (Yu et al., 2024). PGSR combines flattened Gaussians with planarity-aware depth and normal rendering (Chen et al., 2024). Depth, normal, and plane priors provide additional geometric supervision (Turkulainen et al., 2025; Zanjani et al., 2024; Jin et al., 2025; Liu et al., 2025a). These approaches introduce geometric constraints at the pixel, surface, or plane level. However, strictly planar primitives are less flexible when representing irregular geometry.

Existing block-wise methods for urban novel-view synthesis divide large scenes into independently optimized blocks, but this process may result in highly unequal primitive counts across blocks or split connected buildings. Methods using 2DGS effectively represent planar regions in urban scenes, but their zero-thickness primitives are less flexible when representing irregular geometry. UrbanBP-GS addresses these problems through building-aware scene partitioning and planarity-constrained 3D Gaussian updating. Building connectivity is used to refine primitive ownership while limiting the change in each block's primitive count to a predefined budget. Planar instances are then used to encourage orientation consistency among their contributing finite-thickness 3D Gaussians.

3 Method

3.1 Overview

The UrbanBP-GS pipeline is illustrated in Figure 2. Given posed images $\mathcal{I} = \{I_v\}_{v=1}^{V}$ and camera parameters $\{\mathbf{C}_v\}_{v=1}^{V}$, UrbanBP-GS first performs a coarse 3D Gaussian reconstruction to obtain the coarse Gaussian scene $\mathcal{G}^0$. A Gaussian primitive $g_i \in \mathcal{G}^0$ is parameterized as $g_i = (\boldsymbol{\mu}_i, R_i, \mathbf{s}_i, \alpha_i, \mathbf{c}_i)$, where the five terms denote its center, rotation, three optimizable scales, opacity, and appearance coefficients, respectively. The covariance of g_i is $\Sigma_i = R_i \operatorname{diag}(\mathbf{s}_i^2) R_i^\top$, following 3DGS (Kerbl et al., 2023). The three scale components remain optimizable instead of replacing the shortest scale with a zero-thickness dimension. Although no explicit lower bound is imposed on thickness, retaining the third scale preserves the representational freedom needed for irregular geometry.

UrbanBP-GS then applies building-aware scene partitioning to $\mathcal{G}^0$. The primitive centers are projected to bird's-eye view (BEV), where coordinate quantiles initialize blocks with comparable primitive counts. Connected building regions identified from the scene geometry are then used to refine

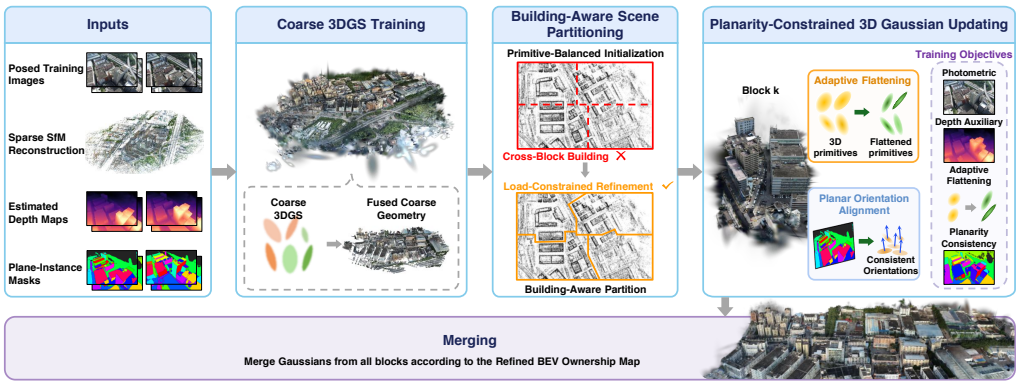

Figure 2: UrbanBP-GS pipeline. A coarse model supports primitive-balanced, building-aware partitioning. Each block then undergoes planarity-constrained 3D Gaussian updating before ownership-guided model merging.

this assignment while keeping changes to each block's primitive count within a predefined budget, producing the refined ownership map Π^*. This map assigns each BEV location to an owning block and is later used to retain the locally trained primitives that belong to that block. Location- and visibility-based view assignment supplies the training observations for the resulting blocks.

Each block is subsequently initialized from the coarse reconstruction and optimized independently through planarity-constrained 3D Gaussian updating. Threshold-controlled Gaussian shaping adjusts each Gaussian until its shortest axis is sufficiently distinct from the other two axes. Plane instances segmented from the training views then identify the Gaussians contributing to each planar region. Soft shortest-axis surface directions are then used to promote orientation consistency within each group without removing the third scale. This stage produces a locally optimized checkpoint for every block. Finally, ownership-guided model merging uses Π^* to remove primitives outside the corresponding ownership region and concatenates the retained Gaussian attributes to obtain the complete urban scene.

3.2 Building-Aware Scene Partitioning

Building-aware scene partitioning proceeds through the four stages illustrated in Figure 3. Primitive-count-balanced initialization first partitions the coarse Gaussian scene and produces an initial ownership map with similar primitive counts. Geometry-derived building identification then extracts connected building regions from the coarse geometry and identifies those spanning multiple blocks. Budget-constrained ownership refinement consolidates each cross-block building into its dominant block when the induced primitive-count variation satisfies the prescribed budget, yielding the refined ownership map Π^*. Finally, view assignment combines camera-location and visibility evidence to construct the training view set for every refined block. By initializing balanced ownership before applying budget-constrained building consolidation, this partitioning procedure reduces building fragmentation while keeping the primitive-count changes of accepted refinements controlled.

Primitive-count-balanced initialization. Each Gaussian primitive $g_i \in \mathcal{G}^0$ is projected onto BEV, producing the coordinate $\mathbf{x}_i \in \mathbb{R}^2$. Coordinate-quantile partitioning constructs an initial ownership map $\Pi^{(0)} : \mathbb{R}^2 \to \{1, \ldots, K\}$. For block k, this map defines the ownership region $P_k^{(0)} = \{\mathbf{x} \in \mathbb{R}^2 \mid \Pi^{(0)}(\mathbf{x}) = k\}$ and the primitive count

$$N_k^{(0)} = \sum_{i=1}^{N} \mathbf{1}\left[\Pi^{(0)}(\mathbf{x}_i) = k\right], \quad \bar{N} = \frac{N}{K}. \tag{1}$$

Here, N is the number of primitives in $\mathcal{G}^0$ and $\bar{N}$ is the target count per block. Candidate grid layouts with different row and column counts are enumerated together with their transposes. A robust working extent is estimated after temporarily excluding extreme primitive centers, and coordinate-quantile cuts are placed within this extent to bring each $N_k^{(0)}$ toward $\bar{N}$. The candidate whose blocks

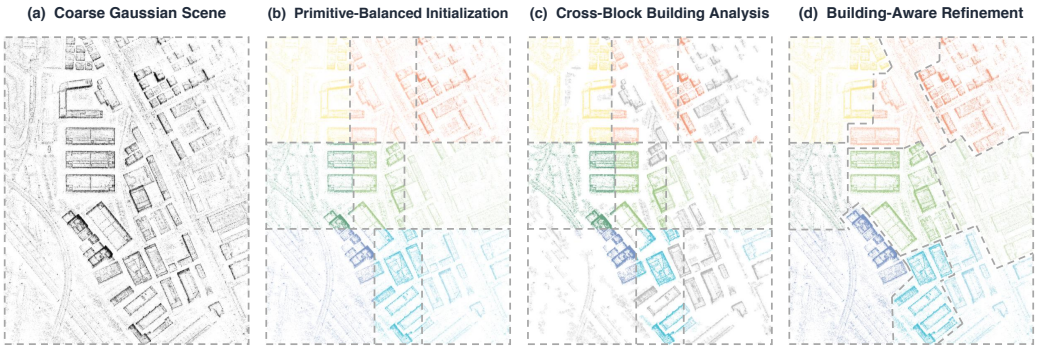

Figure 3: Building-aware scene partitioning in bird's-eye view. The panels show (a) the coarse Gaussian scene, (b) primitive-balanced initialization, (c) cross-block building analysis, and (d) budget-constrained ownership refinement.

have the most compact average aspect ratio is selected. Its outer boundaries are then extended to the complete scene extent, assigning every primitive in $\mathcal{G}^0$ to exactly one block. The resulting map $\Pi^{(0)}$ provides the primitive-count-balanced reference for subsequent building-aware ownership refinement.

Geometry-derived building identification. Geometry-derived building identification extracts connected building regions directly from the coarse Gaussian geometry. A local ground-height map is first estimated, after which coarse Gaussian centers above the local ground are retained and accumulated into a BEV density map. Morphological filtering removes isolated responses. Hole filling and connected-component analysis then produce connected building regions $B_j \subset \mathbb{R}^2$. A region is identified as cross-block when it intersects more than one current ownership region. The candidate owner of building region B_j is the block with the largest BEV overlap, given by $k_j^* = \arg\max_k |B_j \cap P_k|$. This assignment is based on overlap area rather than the building center or a separate count of the primitives associated with B_j. This overlap-based assignment favors the reassignment that changes the smallest portion of the current ownership footprint.

Budget-constrained ownership refinement. Budget-constrained ownership refinement processes the cross-block building regions in descending BEV area. Before region B_j is evaluated, $N_k^{(j-1)}$ denotes the current primitive count in block k, and $\widetilde{N}_k^{(j)}$ denotes the count obtained by assigning all cells of B_j to its dominant block k_j^*. The relative primitive-count variation induced in block k is

$$\Delta_k^{(j)} = \frac{\left|\widetilde{N}_k^{(j)} - N_k^{(j-1)}\right|}{\max\left(N_k^{(j-1)}, 1\right)}. \tag{2}$$

A reassignment is accepted only when $\max_k \Delta_k^{(j)} \leq \delta$. Acceptance immediately updates the current ownership and primitive counts, whereas a region exceeding the budget is skipped. The budget δ therefore controls the primitive-count change of each accepted reassignment relative to the current partition. Its definition does not include cumulative deviation from $\Pi^{(0)}$ or hardware costs. Each accepted reassignment modifies only the rasterized BEV ownership map, leaving the Gaussian parameters unchanged at this stage. Processing all candidate regions produces the refined ownership map Π^*.

View assignment. View assignment associates a training view with a block when its camera center lies in the corresponding base spatial block or when masking that block's coarse Gaussians changes the rendered image beyond a visibility threshold. Combining these location- and visibility-based assignments supplies observations for the nonrectangular regions created by ownership refinement. The refined map determines primitive ownership, while the resulting view set and the common coarse checkpoint initialize independent block optimization.

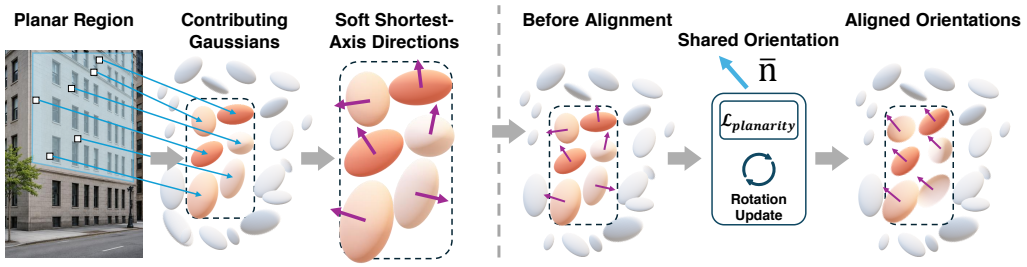

Figure 4: Planarity-constrained Gaussian updating. Contributing Gaussians from a planar region define a shared orientation that guides sign-invariant rotation updates.

3.3 Planarity-Constrained 3D Gaussian Updating

Planarity-constrained 3D Gaussian updating proceeds through the four stages illustrated in Figure 4. Threshold-controlled Gaussian shaping first applies a global scale-ratio regularizer to establish identifiable local short axes while retaining three optimizable scales. Soft surface-direction estimation then converts the rotated local axes into a continuous surface direction $\mathbf{n}_i$ and a scale-derived aggregation weight ω_i. Plane-instance–Gaussian association groups the visible Gaussians contributing to each planar region and produces the set $\mathcal{G}_{vp}$. Planar orientation alignment then estimates a shared orientation for each group and optimizes the Gaussian rotations so that the corresponding surface directions align with it, while opposite directions are treated as equivalent. By combining threshold-controlled shaping with planar orientation alignment, this update procedure exploits planar priors while preserving finite-thickness 3D primitives for irregular geometry.

Threshold-controlled Gaussian shaping. Threshold-controlled Gaussian shaping applies a global scale-ratio regularizer adapted from flattened-Gaussian regularization in PGSR (Chen et al., 2024). For Gaussian i, the ratio $\rho_i = s_i^{\min}/(s_i^{\max} + \epsilon)$ compares its shortest and longest scales. The active set $\mathcal{A} = \{i \mid \rho_i > \tau_s\}$ contains the primitives regulated by the shaping loss, which averages ρ_i over $\mathcal{A}$ and is set to zero when the set is empty. A primitive contributes to this loss only while $\rho_i > \tau_s$. Its shaping gradient ceases once $\rho_i \le \tau_s$. This threshold-controlled operation establishes an identifiable local short axis while retaining three optimizable scale parameters. The threshold affects only the shaping loss and does not set a fixed minimum thickness. A Gaussian is included in $\mathcal{A}$ based only on its scale ratio, regardless of its semantic label or association with a plane instance.

Soft surface-direction estimation. Soft surface-direction estimation extends the shortest-axis construction with a temperature-controlled selector over the three local axes. For Gaussian i, $\mathbf{r}_{im}$ denotes local axis m after rotation into world coordinates, and the selector defines

$$
\begin{aligned}
a_{im} &= \frac{\exp[-\,\mathrm{sg}(s_{im})/T]}{\sum_{n=1}^{3} \exp[-\,\mathrm{sg}(s_{in})/T]}, \\
\mathbf{n}_i &= \mathrm{normalize}\left(\sum_{m=1}^{3} a_{im}\mathbf{r}_{im}\right), \\
\omega_i &= \sum_{m=1}^{3} a_{im} s_{im}.
\end{aligned}
\tag{3}
$$

Here, T controls the selector temperature and sg stops gradients through its scale-dependent weights. The weighted axis combination produces the continuous surface direction $\mathbf{n}_i$ near scale ties and transmits the planarity gradient to Gaussian rotation. Stopping gradients through the selector keeps axis selection fixed during backpropagation, preventing scale updates whose sole purpose is to choose a different local axis. The associated ω_i serves as a scale-derived weight for plane-level aggregation. Alpha-compositing contributions and plane-detector confidences remain separate quantities.

Plane-instance–Gaussian association. Each planar region in a training view is linked to the Gaussians that contribute to its rendering. Plane-instance masks are first extracted from each training view using ZeroPlane (Liu et al., 2025a). For view v, the resulting mask $M_v : \Omega_v \to \{0, \dots, P_v\}$ assigns a plane-instance label to each pixel, with 0 denoting the background. Non-differentiated renderer buffers record the depth-ordered Gaussian IDs contributing to each pixel and the number of valid contributors. For instance p, IDs from pixels satisfying $M_v(\mathbf{u}) = p$ are collected, deduplicated, and filtered by instance size, visibility, and minimum contributor count. The surviving IDs form the associated set $\mathcal{G}_{vp}$. The view index preserves view-specific grouping because plane-instance identities are not matched across views. Membership represents participation in rendering the planar region. Accumulated alpha ranking and unique global plane ownership are outside this association.

Planar orientation alignment. For each planar region, the surface directions of its associated Gaussians are used to estimate a shared orientation. The Gaussian rotations are then optimized so that the corresponding surface directions align with this orientation, while opposite directions are treated as equivalent. Let $\mathcal{P}_v$ denote the valid plane instances in view v. For each (v, p), the scale-derived weights define

$$\begin{aligned} \bar{\mathbf{n}}_{vp} &= \text{normalize}\left(\frac{\sum_{i\in\mathcal{G}_{vp}} \omega_i \mathbf{n}_i}{\sum_{i\in\mathcal{G}_{vp}} \omega_i + \epsilon}\right), \\ \mathcal{L}_{\text{planarity}} &= \frac{1}{|\mathcal{P}_v|} \sum_{p\in\mathcal{P}_v} \frac{\sum_{i\in\mathcal{G}_{vp}} \omega_i \left[1 - (\mathbf{n}_i^\top \bar{\mathbf{n}}_{vp})^2\right]}{\sum_{i\in\mathcal{G}_{vp}} \omega_i + \epsilon}. \end{aligned} \tag{4}$$

The term is set to zero when no instance satisfies the filtering conditions. The squared inner product treats opposite directions as equivalent with respect to the shared orientation. Gradients from this term propagate through $\mathbf{n}_i$ to update the Gaussian rotations. Shortest-scale regulation remains the role of the separate shaping term. The complete update retains the three-scale 3D Gaussian parameterization, preserving finite-thickness primitives without explicit 3D plane fitting or zero-thickness conversion.

3.4 Block Optimization and Model Merging

Block optimization and model merging comprise two stages. Each block is initialized from the common coarse checkpoint and optimized using its assigned training views. The loss terms are introduced progressively during training, producing a local checkpoint for each block. Ownership-guided model merging then filters these checkpoints with Π^* and concatenates the retained attributes into the complete urban scene.

Block optimization. Block optimization initializes each local model from the common coarse checkpoint and applies the 3DGS densification and pruning procedure. Its photometric term is $\mathcal{L}_{\text{rgb}} = 0.8\|\hat{I}_v - I_v\|_1 + 0.2(1 - \text{SSIM}(\hat{I}_v, I_v))$ (Kerbl et al., 2023), where SSIM measures structural similarity (Wang et al., 2004). The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{rgb}} + w_d \mathcal{L}_{\text{depth}} + w_s \mathcal{L}_{\text{shape}} + w_p \mathcal{L}_{\text{planarity}}. \tag{5}$$

Here, w_d, w_s, and w_p are the loss weights for inverse-depth supervision, Gaussian shaping, and planar orientation alignment, respectively. The inverse-depth term is retained from the base training formulation. Gaussian shaping begins after an initial optimization phase. Planar orientation alignment is enabled later, once the shortest axes of the primitives are sufficiently well defined. While the planarity term is active, the renderer buffers rebuild the associations between plane instances and Gaussians for the current training view. The optimized parameters form the local checkpoint used by ownership-guided model merging.

Ownership-guided model merging. Ownership-guided model merging applies Π^* to every local checkpoint after independent training. Gaussians whose centers fall inside the corresponding ownership region are retained, and their positions, scales, rotations, opacities, and appearance attributes are concatenated across blocks.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

The proposed UrbanBP-GS is evaluated on four large-scale urban scenes, including Residence from UrbanScene3D (Lin et al., 2022), Building and Rubble from the Mill 19/Mega-NeRF benchmark (Turki et al., 2022), and MatrixCity-Aerial from MatrixCity (Li et al., 2023). Camera poses follow the provided reconstructions and COLMAP-style preprocessing (Schoenberger & Frahm, 2016). Building, Rubble, and Residence images are downsampled by a factor of four, whereas MatrixCity-Aerial images use a factor of 1.2. Rendering quality is measured by PSNR, SSIM (Wang et al., 2004), and LPIPS (Zhang et al., 2018).

Across all four scenes, the coarse model uses degree-3 spherical harmonics and is trained for 30,000 steps, followed by 60,000 optimization steps for each local block. Residence is partitioned into 12 blocks, Building and Rubble into 9 blocks each, and MatrixCity-Aerial into 16 blocks. Densification is active from step 1,000 to step 30,000. Threshold-controlled Gaussian shaping starts at step 3,000 with $\tau_s = 0.2$, $T = 0.1$, and a scheduler initialized at 0.01, while planarity consistency starts at step 6,000 with a scheduler initialized at 0.2. All scenes use a 1024×1024 ownership grid, a per-refinement primitive-count variation budget of 0.3, and a view-visibility threshold of 0.05. The inherited monocular inverse-depth schedule is held fixed across datasets and ablation variants.

4.2 COMPARISON STUDY

We compare UrbanBP-GS with the NeRF-based methods Mega-NeRF (Turki et al., 2022) and Switch-NeRF (Mi & Xu, 2023), as well as the Gaussian-splatting-based methods VastGaussian (Lin et al., 2024), CityGaussian (Liu et al., 2024), CityGaussianV2 (Liu et al., 2025b), BlockGaussian (Wu et al., 2025), and LoBE-GS (Hung et al., 2025). As shown in Table 1, UrbanBP-GS achieves the lowest LPIPS on all four scenes and the highest PSNR on Building, Rubble, and Residence. On Residence, it performs best across all three metrics, improving PSNR by 0.75 dB and reducing LPIPS from 0.209 to 0.184 relative to the best baseline for each metric. On MatrixCity-Aerial, its PSNR is lower than that of VastGaussian, but it achieves the lowest LPIPS of 0.139, compared with 0.169 for CityGaussianV2, indicating better perceptual quality.

Table 1: Novel-view synthesis on four large-scale urban scenes.

	Building			Rubble			Residence			MatrixCity-Aerial		
Method	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Mega-NeRF (CVPR 2022)	20.92	0.547	0.454	24.06	0.553	0.508	22.08	0.628	0.401	-	-	-
Switch-NeRF (ICLR 2023)	21.54	0.579	0.397	24.31	0.562	0.478	22.57	0.654	0.352	-	-	-
VastGaussian (CVPR 2024)	21.80	0.728	0.225	25.20	0.742	0.264	21.01	0.699	0.261	**28.33**	0.835	0.220
CityGaussian (ECCV 2024)	21.55	0.778	0.246	25.77	**0.813**	0.228	22.00	0.813	0.211	27.46	0.865	0.204
CityGaussianV2 (ICLR 2025)	19.07	0.650	0.397	23.75	0.720	0.322	21.15	0.769	0.234	27.23	0.857	0.169
BlockGaussian (arXiv 2025)	21.11	0.750	0.234	25.52	0.801	0.233	22.15	0.810	0.211	28.17	0.863	0.199
LoBE-GS (arXiv v2, 2026)	22.13	**0.790**	0.231	25.93	**0.813**	0.232	21.48	0.805	0.209	27.69	**0.875**	0.185
UrbanBP-GS (Ours)	**22.63**	0.787	**0.195**	**26.09**	0.803	**0.203**	**23.32**	**0.828**	**0.184**	27.44	0.871	**0.139**

Figure 5 shows visual comparisons with CityGaussianV2 and BlockGaussian. In the highlighted regions, UrbanBP-GS produces clearer facade boundaries and more consistent planar textures. It also better preserves the crane's slender elements and distinct silhouette against the surrounding background. Overall, these results show competitive quantitative performance and improved visual quality.

4.3 ABLATION STUDY

We conduct remove-one ablations on Residence and use the same training settings for all variants. The evaluation retains PSNR, SSIM, and LPIPS to remain aligned with the novel-view-synthesis scope.

Removing building-aware refinement degrades all three rendering metrics on Residence, with the largest displayed changes in SSIM and LPIPS among the evaluated variants. These results suggest



Figure 5: Qualitative comparison with CityGaussianV2 and BlockGaussian.

that assigning connected parts of a building to the same block helps reduce inconsistencies caused by independent block optimization.

Removing planarity consistency produces a small but unfavorable shift in all three underlying measurements. The modest magnitude is consistent with the targeted role of the objective, which coordinates orientations only among Gaussians associated with valid planar regions. In contrast, the reported image-space metrics are averaged over entire views containing sky, vegetation, and other nonplanar content.

Table 2: Remove-one ablation on Residence.

Variant	PSNR ↑	SSIM ↑	LPIPS ↓
Full	**23.315**	**0.828**	**0.184**
w/o Building-Aware	23.150	0.824	0.187
w/o Planarity Consistency	23.188	0.828	0.185
w/o Gaussian Shaping	23.031	0.825	0.185

Performance drops more when threshold-controlled Gaussian shaping is removed than when planar orientation alignment is removed. This trend is consistent with Gaussian shaping providing a more stable short-axis basis for soft surface-direction estimation within planarity-constrained updating.

5 Conclusion

We propose UrbanBP-GS for block-wise novel-view synthesis of large-scale urban scenes, addressing limitations in scene partitioning and Gaussian primitive representation. Building-aware scene partitioning refines a primitive-balanced ownership map using geometry-derived connected-building regions under an incremental primitive-count variation budget. Planarity-constrained updating associates detected planar regions with contributing 3D Gaussians and coordinates their surface directions while retaining the flexibility of a three-scale 3D representation. Across four evaluated scenes, UrbanBP-GS obtains the lowest LPIPS on all four and the highest PSNR on three, while the Residence ablation supports the rendering contributions of both core components.

AI USE STATEMENT

Generative AI tools were used for language editing and to assist with editing, implementing, and debugging the research code. The authors reviewed and verified all AI-assisted outputs and take full responsibility for the final manuscript and implementation.

REFERENCES

Danpeng Chen, Hai Li, Weicai Ye, Yifan Wang, Weijian Xie, Shangjin Zhai, Nan Wang, Haomin Liu, Hujun Bao, and Guofeng Zhang. Pgsr: Planar-based gaussian splatting for efficient and high-fidelity surface reconstruction. *arXiv preprint arXiv:2406.06521*, 2024. URL https://arxiv.org/abs/2406.06521.

Yu Chen and Gim Hee Lee. Dogs: Distributed-oriented gaussian splatting for large-scale 3d reconstruction via gaussian consensus. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/3d03b03197666b19c6a6e69812dd3e34-Abstract-Conference.html.

Antoine Guedon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Guedon_SuGaR_Surface-Aligned_Gaussian_Splatting_for_Efficient_3D_Mesh_Reconstruction_and_CVPR_2024_paper.html.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. *ACM Transactions on Graphics*, 2024. URL https://arxiv.org/abs/2403.17888.

Sheng-Hsiang Hung, Ting-Yu Yen, Wei-Fang Sun, Simon See, Shih-Hsuan Hung, and Hung-Kuo Chu. Lobe-gs: Load-balanced and efficient 3d gaussian splatting for large-scale scene reconstruction. *arXiv preprint arXiv:2510.01767*, 2025. URL https://arxiv.org/abs/2510.01767v2. Version 2, revised April 2026.

Xirui Jin, Renbiao Jin, Boying Li, Danping Zou, and Wenxian Yu. Planargs: High-fidelity indoor 3d gaussian splatting guided by vision-language planar priors. In *Advances in Neural Information Processing Systems*, 2025. URL https://papers.neurips.cc/paper_files/paper/2025/hash/33c674cb3ce9dae35021930d8d63308f-Abstract-Conference.html.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

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 *IEEE/CVF International Conference on Computer Vision*, pp. 3205–3215, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Li_MatrixCity_A_Large-scale_City_Dataset_for_City-scale_Neural_Rendering_and_ICCV_2023_paper.html.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. URL https://arxiv.org/abs/2402.17427.

Liqiang Lin, Yilin Liu, Yue Hu, Xingguang Yan, Ke Xie, and Hui Huang. Capturing, reconstructing, and simulating: The urbanscene3d dataset. In *European Conference on Computer Vision*, 2022. URL https://arxiv.org/abs/2107.04286.

Jiachen Liu, Rui Yu, Sili Chen, Sharon X. Huang, and Hengkai Guo. Towards in-the-wild 3d plane reconstruction from a single image. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025a. URL https://openaccess.thecvf.com/content/CVPR2025/html/Liu_Towards_In-the-wild_3D_Plane_Reconstruction_from_a_Single_Image_CVPR_2025_paper.html.

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*, 2024. URL https://arxiv.org/abs/2404.01133.

Yang Liu, Chuanchen Luo, Zhongkai Mao, Junran Peng, and Zhaoxiang Zhang. Citygaussianv2: Efficient and geometrically accurate reconstruction for large-scale scenes. In *International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=a3ptUbuzbW.

Zhenxing Mi and Dan Xu. Switch-nerf: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PQ2zoIZqvm.

Johannes L. Schoenberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. URL https://openaccess.thecvf.com/content_cvpr_2016/html/Schonberger_Structure-From-Motion_Revisited_CVPR_2016_paper.html.

Yawar Siddiqui, Lorenzo Porzi, Samuel Rota Bulò, Norman Müller, Matthias Nießner, Angela Dai, and Peter Kontschieder. Panoptic lifting for 3d scene understanding with neural fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9043–9052, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Siddiqui_Panoptic_Lifting_for_3D_Scene_Understanding_With_Neural_Fields_CVPR_2023_paper.html.

Mai Su, Zhongtao Wang, Huishan Au, Yilong Li, Xizhe Cao, Chengwei Pan, Yisong Chen, and Guoping Wang. Hug: Hierarchical urban gaussian splatting with block-based reconstruction for large-scale aerial scenes. In *IEEE/CVF International Conference on Computer Vision*, pp. 28839–28848, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Su_HUG_Hierarchical_Urban_Gaussian_Splatting_with_Block-Based_Reconstruction_for_Large-Scale_ICCV_2025_paper.html.

Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P. Srinivasan, Jonathan T. Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. URL https://arxiv.org/abs/2202.05263.

Haithem Turki, Deva Ramanan, and Mahadev Satyanarayanan. Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. URL https://arxiv.org/abs/2112.10703.

Haithem Turki, Jason Y. Zhang, Francesco Ferroni, and Deva Ramanan. Suds: Scalable urban dynamic scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12375–12385, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Turki_SUDS_Scalable_Urban_Dynamic_Scenes_CVPR_2023_paper.html.

Matias Turkulainen, Xuqian Ren, Iaroslav Melekhov, Otto Seiskari, Esa Rahtu, and Juho Kannala. Dn-splatter: Depth and normal priors for gaussian splatting and meshing. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2421–2431, 2025. URL https://openaccess.thecvf.com/content/WACV2025/html/Turkulainen_DN-Splatter_Depth_and_Normal_Priors_for_Gaussian_Splatting_and_Meshing_WACV_2025_paper.html.

Mengyi Wang, Beiqi Chen, Shengfang Pan, Niansheng Liu, and Jinhe Su. Consistency-preserving gaussian splatting for block-based large-scale scene reconstruction. *Computers & Graphics*, 134:

104493, 2026. doi: 10.1016/j.cag.2025.104493. URL https://doi.org/10.1016/j.cag.2025.104493.

Tengfei Wang, Xin Wang, Yongmao Hou, Yiwei Xu, Wendi Zhang, and Zongqian Zhan. Pgsag: Parallel gaussian splatting for fine-grained large-scale urban buildings reconstruction via semantic-aware grouping. *arXiv preprint arXiv:2501.01677*, 2025. URL https://arxiv.org/abs/2501.01677.

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. URL https://doi.org/10.1109/TIP.2003.819861.

Yongchang Wu, Zipeng Qi, Zhenwei Shi, and Zhengxia Zou. Blockgaussian: Efficient large-scale scene novel view synthesis via adaptive block-based gaussian splatting. *arXiv preprint arXiv:2504.09048*, 2025. URL https://arxiv.org/abs/2504.09048.

Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics*, 2024. URL https://arxiv.org/abs/2404.10772.

Farhad G. Zanjani, Hong Cai, Hanno Ackermann, Leila Mirvakhabova, and Fatih Porikli. Planar gaussian splatting. *arXiv preprint arXiv:2412.01931*, 2024. URL https://arxiv.org/abs/2412.01931.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. URL https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.