OBSGS: OBSERVABILITY-GATED DENSIFICATION FOR LARGE-SCENE GAUSSIAN RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Large-scene Gaussian reconstruction must spend a limited primitive budget on centres the cameras can triangulate. Standard adaptive density control instead clones or splits where the view-space photometric gradient is large. On a forward city road the two decisions diverge: many cameras share one bearing, the image residual can still be large, and the depth of a new centre stays weakly determined. We introduce ObsGS, which multiplies that photometric score by the scale-normalized smallest eigenvalue of the multiview projection information matrix. The renderer, photometric loss, and optimizer are unchanged, and training does not read depth. On two MatrixCity street acquisitions, with ground-truth cameras, a shared $2{,}000$-step schedule, and a matched cap of $10{,}000$ Gaussians, ObsGS raises centre-to-surface F-score at $2\,\mathrm{m}$ from 0.099 to 0.143 on the road and from 0.245 to 0.293 on the vertical sequence, above a view-count gate in both cases. The advantage is a camera-only control that moves a fixed budget toward recoverable surfaces. A bearing-angle gate captures much of the same surface movement, and on the vertical sequence ObsGS lowers held-out PSNR by $2.32\,\mathrm{dB}$, so the measured gain is directional observability with a bounded image-quality cost.

1 INTRODUCTION

Neural radiance fields made novel-view synthesis a standard scene representation, including unbounded outdoor capture (Mildenhall et al., 2020; Barron et al., 2022; Müller et al., 2022). City-scale 3D Gaussian Splatting now trains the same problem with explicit primitives, through partitioning, level-of-detail, and distributed optimization (Kerbl et al., 2023; Lin et al., 2024; Liu et al., 2024; 2025; Kerbl et al., 2024). Those systems decide how a large model is executed. They still create primitives with the rule inherited from object-scale splatting: a Gaussian is cloned or split when its accumulated view-space positional gradient is large (Kerbl et al., 2023). That gradient measures photometric demand. In a large street scene it is a weak proxy for geometric recoverability. A forward-looking road supplies many frames and almost one viewing direction, so a point near the motion axis can attract a large image residual while its depth remains poorly constrained. Under an explicit primitive budget, each such split displaces a centre that could have been spent on a facade, a corner, or another surface the cameras actually triangulate.

The recent literature has pushed scale, surface representation, and gradient bookkeeping, and it has left this local decision open. VastGaussian partitions a scene and merges block models (Lin et al., 2024). CityGaussian and CityGaussianV2 add urban level-of-detail, with the latter also changing densification and filtering toward geometry and adding depth-aware supervision (Liu et al., 2024; 2025). 2D Gaussian Splatting replaces ellipsoids with disks and already notes that densification can follow texture (Huang et al., 2024). AbsGS repairs cancellation in the positional gradient so that fine appearance is not missed (Ye et al., 2024). PUP 3D-GS prunes a finished model with a photometric uncertainty score (Hanson et al., 2025). These changes either assume a density rule inside each block, alter the primitive and the loss, repair the image-space signal, or act after training. None of them asks, before the split and at a matched count, whether the training cameras determine the new centre. View count does not close the gap: collinear cameras still increment the count.

ObsGS is a densification gate for that setting. At each densification step it builds, from the training cameras and the current centres alone, the 3×3 Gram matrix of the perspective-projection Jacobian,

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.

converts the smallest eigenvalue into a scale-free score, and multiplies the standard photometric score by a quantile gate. Figure 1 shows where this gate sits: the renderer, the photometric loss, and the optimizer stay on their original path, and every compared rule shares the same primitive cap. The contribution is threefold.

- We formulate budgeted densification in a large scene as the product of photometric demand and multiview recoverability, and we identify collinear street cameras as the case in which a large view count still leaves a zero smallest eigenvalue.
- We derive a scale-normalized weakest-eigenvalue gate from the projection information matrix, prove that it lies in $[0, 1]$ and is invariant to a positive rescaling of that matrix, and insert it without depth supervision or a change of renderer.
- On two MatrixCity acquisitions at a matched budget of about $10{,}000$ Gaussians, the gate improves $2\,\mathrm{m}$ surface F-score over standard densification and over view count, and the comparison with a bearing-angle gate localizes that improvement to directional observability.

2 Related work

Scalable Gaussian reconstruction decides where computation happens. Block partition, hierarchies, level-of-detail, and distributed training make a city representable (Lin et al., 2024; Liu et al., 2024; 2025; Kerbl et al., 2024; Lu et al., 2024; Ren et al., 2025; Chen & Lee, 2024; Zhao et al., 2025; Gao et al., 2025; Li et al., 2025). Inside each executable block the primitive is still created by a local rule. ObsGS replaces that rule and leaves the partition, the rasterizer, and the photometric objective in place, so it is complementary to a city-scale trainer rather than a substitute for one.

Surface-oriented splatting changes the primitive or adds a geometric loss. 2D Gaussian Splatting and SuGaR align kernels with surfaces, and Mip-Splatting controls aliasing as the footprint changes with depth (Huang et al., 2024; Guédon & Lepetit, 2024; Yu et al., 2024). CityGaussianV2 continues the geometric line for urban scenes (Liu et al., 2025). A gain from those systems mixes a new representation, a new loss, and a new density heuristic. Our comparison freezes the 3D Gaussian renderer and the photometric loss, so a change in the surface metric can be attributed to the decision to create primitives.

Other densification scores stay inside the rendering loss. AbsGS accumulates absolute positional gradients to avoid cancellation (Ye et al., 2024). PUP 3D-GS uses a Fisher-style approximation to prune uncertain primitives after training (Hanson et al., 2025). Levenberg–Marquardt splatting changes the optimizer rather than the split criterion (Höllein et al., 2025). ObsGS is applied before the split, uses the camera Jacobian rather than the Hessian of pixel error, and keeps the primitive when the gate is only moderately small. The experiments therefore compare creation rules at matched count. They do not claim a compression result.

The gate assumes calibrated cameras. Classical structure from motion and bundle adjustment, and their global variants, are what supply that calibration in the field (Schönberger & Frahm, 2016; Triggs et al., 2000; Pan et al., 2024). Feed-forward geometry models now estimate points and cameras from images alone (Wang et al., 2024; Leroy et al., 2024; Wang et al., 2025; Yang et al., 2025). This paper uses the dataset poses and a COLMAP initialization so that the densification comparison is not confounded by pose error.

3 Method

3.1 Budgeted densification

Let Gaussian j have centre $\boldsymbol{\mu}_j \in \mathbb{R}^3$. Training cameras are calibrated pinholes. The renderer, the photometric loss $\mathcal{L}_{\mathrm{rgb}}$, and the optimizer are those of 3D Gaussian Splatting (Kerbl et al., 2023). Standard adaptive density control clones or splits when the view-space score

$$s_j^{\mathrm{rgb}} = \left\| \frac{1}{|\mathcal{T}_j|} \sum_{i \in \mathcal{T}_j} \nabla_{\mathbf{u}_{ij}} \mathcal{L}_{\mathrm{rgb}} \right\|_2 \tag{1}$$

exceeds a threshold, subject to the usual scale and opacity tests. $\mathbf{u}_{ij}$ is the projected mean and $\mathcal{T}_j$ indexes the iterations in which the Gaussian is visible. We do not change the estimator of s_j^{rgb}. We multiply it by a gate $g_j \in [g_{\min}, 1]$ computed from the cameras and the current centres. The budget is a hard cap: after each densification step the lowest opacities are pruned until at most $10{,}000$ Gaussians remain. Every baseline in Section 4 uses this cap, so a higher surface score is a reallocation of the same budget.

3.2 Projection information

Write the world-to-camera map of camera i as $\mathbf{p}_{ij} = \mathbf{R}_i \boldsymbol{\mu}_j + \mathbf{t}_i = (x, y, z)^\top$ with $z > 0$, and the pinhole map, up to the principal point, as

$$\boldsymbol{\pi}_i(\mathbf{p}) = \begin{pmatrix} f_{x,i}\, x/z \\ f_{y,i}\, y/z \end{pmatrix}. \tag{2}$$

Differentiating at $\mathbf{p}_{ij}$ gives the camera-frame Jacobian

$$\mathbf{D}_i(\mathbf{p}_{ij}) = \begin{bmatrix} f_{x,i}/z & 0 & -f_{x,i}x/z^2 \\ 0 & f_{y,i}/z & -f_{y,i}y/z^2 \end{bmatrix}. \tag{3}$$

The chain rule $\partial \mathbf{p}_{ij} / \partial \boldsymbol{\mu}_j = \mathbf{R}_i$ yields the world Jacobian $\mathbf{J}_{ij} = \mathbf{D}_i(\mathbf{p}_{ij})\, \mathbf{R}_i$. Over training cameras whose frustum contains the centre, the projection information is the Gram matrix

$$\mathbf{H}_j = \sum_{i \in \mathcal{V}_j} \mathbf{J}_{ij}^\top \mathbf{J}_{ij}. \tag{4}$$

Weights are uniform. Occlusion, opacity, and pose covariance are not included, and visibility is the geometric frustum. Equation equation 4 is the Gauss–Newton matrix of an isotropic projection residual. It is not the Hessian of $\mathcal{L}_{\mathrm{rgb}}$.

One camera cannot determine a centre. For any $\lambda > 0$, $\boldsymbol{\pi}_i(\lambda \mathbf{p}) = \boldsymbol{\pi}_i(\mathbf{p})$, so the derivative of $\boldsymbol{\pi}_i$ along $\mathbf{p}$ vanishes and $\mathbf{D}_i(\mathbf{p})\, \mathbf{p} = \mathbf{0}$. Thus $\mathbf{J}_{ij}^\top \mathbf{J}_{ij}$ has rank at most two and a zero eigenvalue, with null direction along the ray from the camera centre through $\boldsymbol{\mu}_j$. Summing cameras preserves a null direction only when that direction is shared. If $\mathbf{v} \neq \mathbf{0}$ satisfies $\mathbf{v}^\top \mathbf{H}_j \mathbf{v} = 0$, then $\mathbf{J}_{ij}\mathbf{v} = \mathbf{0}$ for every $i \in \mathcal{V}_j$, so $\mathbf{v}$ is parallel to every viewing ray. On a forward road, a centre near the motion axis has a large view count and a zero smallest eigenvalue at once. That is the large-scene failure view count cannot see. The supplementary material records the short proofs and the case in which two non-parallel rays make $\mathbf{H}_j$ positive definite.

3.3 Scale-normalized gate

The absolute eigenvalue $\lambda_{\min}(\mathbf{H}_j)$ shrinks as z grows, because the entries of $\mathbf{D}_i$ scale as f/z and f/z^2. A near, poorly angled point can therefore outrank a far, well angled one. We score the shape of $\mathbf{H}_j$ instead,

$$o_j = \frac{\lambda_{\min}(\mathbf{H}_j)}{\mathrm{tr}(\mathbf{H}_j)/3 + \epsilon}. \tag{5}$$

Proposition 1. *If $\mathbf{H}$ is symmetric positive semidefinite and $c > 0$, then $o(c\mathbf{H}) = o(\mathbf{H})$. If $\mathrm{tr}(\mathbf{H}) > 0$ and $\epsilon = 0$, then $o(\mathbf{H}) \in [0, 1]$, and $o(\mathbf{H}) = 1$ only when $\mathbf{H}$ is a positive multiple of the identity.*

The first claim is immediate because every eigenvalue, and the trace, scales by c. The second holds because $\lambda_{\min}$ is nonnegative for a Gram matrix and cannot exceed the mean eigenvalue $\mathrm{tr}(\mathbf{H})/3$. The supplementary material writes both arguments in full. Among centres seen by at least two cameras, let Q_q be the empirical q-quantile of o_j. The gate is

$$g_j = \begin{cases} g_{\min} + (1 - g_{\min})\, \mathrm{clip}(o_j / Q_q,\, 0,\, 1), & |\mathcal{V}_j| \geq 2, \\ g_{\min}, & \text{otherwise}, \end{cases} \tag{6}$$

and the densification score is $s_j = g_j s_j^{\mathrm{rgb}}$. The default is $q = 0.5$ and $g_{\min} = 0.05$. These values are the implementation default. They were not tuned on the vertical acquisition. Centres with fewer

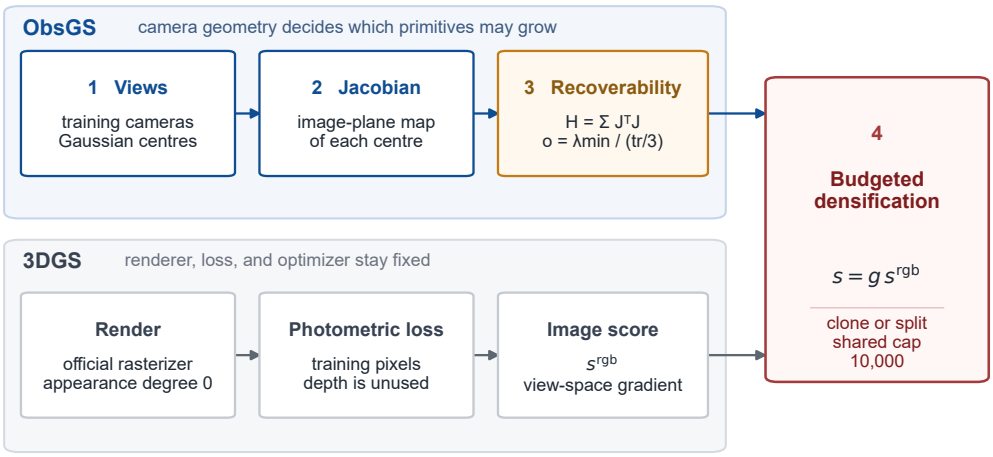


Figure 1: ObsGS leaves the 3D Gaussian Splatting loop unchanged and multiplies its view-space score by a recoverability gate. Training cameras and current centres form the projection information $\mathbf{H}$. The scale-normalized smallest eigenvalue becomes a quantile gate g. Clone and split then use $s = g\, s^{\mathrm{rgb}}$ under one shared cap. View count and bearing angle replace the score inside the same gate.

than two views receive the floor because a single ray has $\lambda_{\min} = 0$ identically, so their score carries no ranking information. The map $o \mapsto g$ is nondecreasing, and $g_j \in [g_{\min}, 1]$.

Two controls enter Equation equation 6 with a different scalar. View count uses $|\mathcal{V}_j|$. The angle control replaces $\mathbf{H}_j$ by $\sum_i (\mathbf{I} - \mathbf{d}_{ij}\mathbf{d}_{ij}^\top)$, where $\mathbf{d}_{ij}$ is the unit bearing from the camera centre to the Gaussian, and again reads $\lambda_{\min}$. Each summand annihilates its own bearing and equals the identity on the orthogonal plane, so it has the same kernel condition as $\mathbf{J}_{ij}^\top \mathbf{J}_{ij}$ and none of the depth or focal-length weighting. The experiments measure what that weighting adds. Absolute $\lambda_{\min}$, trace, and log-determinant are derived in the supplement and are not the default rule.

The score uses training cameras and current centres at densification steps only. It does not read ground-truth depth, test images, or test poses. Depth appears only in the surface evaluation of Section 4.

4 Experiments

4.1 Environment, Data, and Protocol

All runs used a Linux server with NVIDIA RTX A6000 GPUs (48 GB). Each job occupied one GPU. The implementation is PyTorch 2.11.0 with CUDA 12.8 and the official 3D Gaussian Splatting CUDA rasterizer. Under the schedule below, wall time was 9–15 seconds and the modal allocated memory was about 217 MiB. The cap, rather than the device memory, is what matches the budget.

The data are two MatrixCity street acquisitions (Li et al., 2023): a forward road trajectory (Down) and a vertical-looking trajectory (Vertical). Each run uses 64 training images and 8 held-out images. Cameras are the dataset's ground-truth poses. The initial point cloud is a 2,500-point COLMAP reconstruction (Schönberger & Frahm, 2016) aligned to that pose gauge, so the study isolates densification with known cameras. It is not a pose-free reconstruction. Images are optimized for 2,000 steps at 384^2 with spherical harmonics of degree 0. Densification runs from step 500 to 1,500 every 100 steps, at the vendor gradient threshold 2×10^{-4}. The cap of 10,000 keeps the highest opacities. Seeds are 0, 1, and 2. Final counts lie between 9,909 and 9,985.

Held-out PSNR and SSIM use the 8 test views. Surface metrics use a different set of frames: Gaussian centres with opacity at least 0.05 are compared with back-projected metric depth on training-trajectory frames $0, 8, \ldots, 56$. Accuracy is the distance from a centre to the observed surface. Com-

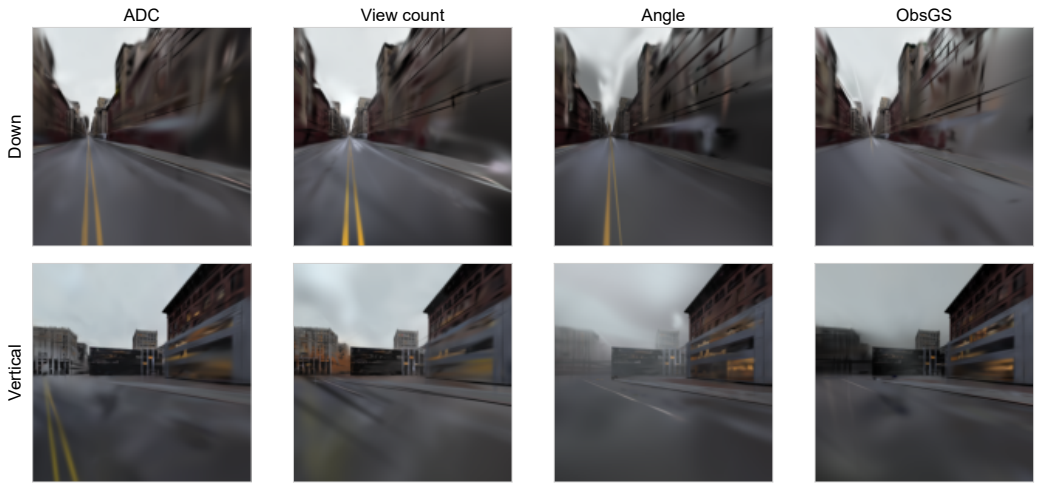


Figure 2: Held-out renders from the saved $2{,}000$-step models, seed 0, rasterized at 768^2. In each acquisition the camera is the test view on which standard densification has the highest PSNR. Top: Down. Bottom: Vertical. Columns share initialization, that camera, and the $10{,}000$ cap. Table 1 still averages all eight test views at the training resolution of 384^2.

pleteness is the distance from a surface sample to the nearest centre. F-score is the harmonic mean of the fractions within a stated radius. We report F-score at $2\,\mathrm{m}$ as the primary surface number because mean distances are heavy-tailed, and we also report median completeness. Depth is evaluation-only. The question is whether $g_j s_j^{\mathrm{rgb}}$ improves this surface score, at matched count, over the image gradient, view count, and bearing angle, and what image quality the reallocation costs. We treat a held-out PSNR drop above $0.3\,\mathrm{dB}$ relative to standard densification as a material image-quality cost.

4.2 Main comparison

Table 1 gives three-seed means. On Down, ObsGS raises F-score at $2\,\mathrm{m}$ from 0.099 to 0.143 and lowers mean centre-to-surface accuracy from $87.4\,\mathrm{m}$ to $50.1\,\mathrm{m}$. View count reaches 0.122. The angle gate reaches 0.153, with PSNR $13.07\,\mathrm{dB}$ against $12.88\,\mathrm{dB}$ for standard densification and $12.34\,\mathrm{dB}$ for ObsGS. Median completeness is $5.81\,\mathrm{m}$ for ObsGS and $5.92\,\mathrm{m}$ for angle. On Vertical, ObsGS has the highest F-score, 0.293 against 0.288 for angle, 0.252 for view count, and 0.245 for standard densification. Mean accuracy falls from $38.6\,\mathrm{m}$ to $15.1\,\mathrm{m}$. Held-out PSNR falls from $18.16\,\mathrm{dB}$ to $15.84\,\mathrm{dB}$, a $2.32\,\mathrm{dB}$ drop, while angle falls to $16.39\,\mathrm{dB}$ and reaches a better mean accuracy ($10.0\,\mathrm{m}$). Mean completeness stays above $65\,\mathrm{m}$ for every rule. The median is the stable completeness summary.

Figure 2 shows the held-out frames for seed 0. At this budget the road renders remain incomplete. The vertical renders differ in facade continuity and floaters, consistent with a reallocation of the same primitive count rather than with a uniformly sharper reconstruction. Figure 3 plots the three-seed means. The same ordering, with seed-to-seed changes in floaters and road detail, appears for seeds 1 and 2 in the supplementary material. ObsGS is the only rule that improves F-score over both standard densification and view count on both acquisitions. The angle gate shows that directional diversity, without the focal-length weighting inside $\mathbf{J}$, accounts for most of that surface movement. The full Jacobian adds a small Vertical F-score edge and does not win the road F-score or the PSNR comparison.

4.3 Ablations

Table 2 varies one default at a time. On Down, $q = 0.75$ reaches F-score 0.147, still below angle, and $g_{\min} = 0$ matches the default F-score while recovering PSNR to $12.90\,\mathrm{dB}$. Raising the floor to 0.2 lowers F-score toward the view-count regime: the gate then stops suppressing weak directions. Camera stride 4, which subsamples the training cameras inside the score only, stays at F-score 0.126.

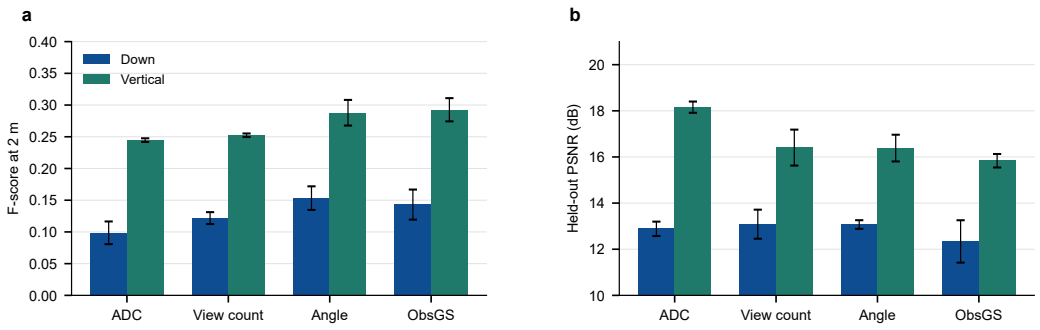


Figure 3: Three-seed comparison. Bars are means and whiskers are standard deviations across seeds 0, 1, and 2. **(a)** Centre-to-surface F-score at 2 m. **(b)** Held-out PSNR. ObsGS improves F-score over standard densification and view count on both acquisitions.

Table 1: Matched-count densification on MatrixCity. Entries are means over three seeds. Accuracy and completeness are in metres. About 9,900–9,985 Gaussians remain under the 10,000 cap.

Scene	Rule	Acc. ↓	Comp. med. ↓	F@2 m ↑	PSNR ↑	SSIM ↑
Down	Standard ADC	87.41	7.55	0.099±0.018	12.88±0.31	0.358
Down	View count	74.05	6.42	0.122±0.009	13.09±0.63	0.376
Down	Angle	55.57	5.92	**0.153**±0.019	**13.07**±0.19	0.395
Down	ObsGS	**50.13**	**5.81**	0.143±0.024	12.34±0.92	0.388
Vertical	Standard ADC	38.61	11.74	0.245±0.003	**18.16**±0.24	**0.632**
Vertical	View count	28.15	11.02	0.252±0.003	16.41±0.78	0.609
Vertical	Angle	**10.03**	**9.28**	0.288±0.020	16.39±0.58	0.620
Vertical	ObsGS	15.13	9.81	**0.293**±0.018	15.84±0.29	0.584

On Vertical, the listed departures from the default lower F-score, and $g_{\min} = 0$ drops PSNR further, to 14.31 dB. The gate is not highly sensitive to the exact quantile, and it does not require every camera. No ablation here reverses the road comparison with angle or brings Vertical PSNR back within 0.3 dB of standard densification. Absolute $\lambda_{\min}$ on Down seed 0 reached F-score 0.198 with PSNR 12.11 dB and mean completeness 45.0 m, worse than the three-seed standard-densification mean completeness of 24.7 m. We treat that single seed as evidence of the near-depth bias predicted by the $1/z$ scaling, not as a selected operating point. Further stride settings and a non-rendering allocation check are in the supplement.

4.4 What the budget comparison shows

Final counts differ by less than one percent, so the F-score gaps are not an artifact of using fewer primitives. Because the photometric loss is shared, the Vertical PSNR drop is evidence that suppressing low-observability splits changes which image error the optimizer can fit. The gate does what it was built to do: it spends the same count on a different set of centres. In this scene pair, angular diversity explains most of the surface gain, and the image-quality cost on the vertical acquisition is part of the result. The gate does not create the missing observations. A city method that adds depth or normal supervision (Liu et al., 2025) answers a different question, and this paper does not run that comparison.

5 Conclusion

ObsGS is a camera-only densification gate for large-scene Gaussian reconstruction under a hard primitive budget. It multiplies the standard photometric score by the normalized smallest eigenvalue of a multiview projection information matrix, leaves rendering and the loss unchanged, and improves 2 m surface F-score over standard densification and over view count on both a forward road and a vertical street acquisition from MatrixCity. The comparison with a bearing gate shows that the

Table 2: ObsGS ablations. Means over three seeds unless noted. The default is $q = 0.5$, $g_{\min} = 0.05$, and camera stride 1.

Scene	Setting	F@2 m ↑	PSNR ↑	Acc. ↓	Comp. med. ↓
Down	Default	0.143	12.34	50.13	5.81
Down	$q = 0.25$	0.126	12.75	55.57	6.39
Down	$q = 0.75$	0.147	12.75	47.10	5.72
Down	$g_{\min} = 0$	0.143	12.90	44.63	5.95
Down	$g_{\min} = 0.2$	0.121	13.22	62.83	6.96
Down	Stride 4	0.126	13.16	57.81	7.18
Vertical	Default	0.293	15.84	15.13	9.81
Vertical	$q = 0.75$ (seed 0)	0.328	16.58	8.73	5.47
Vertical	$g_{\min} = 0$	0.284	14.31	13.05	8.83
Vertical	Stride 4	0.272	15.93	11.74	11.64
Vertical	Stride 8	0.260	16.21	9.83	11.89

operative geometric factor is directional diversity. That is a concrete contribution to how city-scale systems should spend a local budget: collinear views are not coverage, and a Jacobian gate makes that fact a number the densifier can use. The same measurements bound the claim. The full projection weighting does not outperform bearings on the road F-score, and on the vertical acquisition the PSNR drop exceeds $0.3\,\mathrm{dB}$.

The next step for large-scene reconstruction is to carry this decision into the trainers that already partition a city. Inside each block the densifier is still local, so the same gate can sit under a level-of-detail hierarchy without a new renderer. The open technical questions are the ones this protocol held fixed: visibility from rasterized transmittance rather than from the frustum, cameras estimated rather than given, schedules long enough for high-order appearance, and streets captured in the field rather than rendered from MatrixCity. Those extensions test whether directional observability remains the right local prior once the block, the pose, and the capture noise are no longer ideal. They are the path from a matched-budget gate to a city system that spends primitives where the surfaces are recoverable.

AI USE STATEMENT

We used generative AI tools to assist with drafting, restructuring, and language editing of the manuscript. The authors reviewed the text and are responsible for the final content.

REPRODUCIBILITY STATEMENT

Section 4 states the hardware, software, scenes, seeds, camera source, initialization, schedule, threshold, cap, and the split between held-out PSNR and training-trajectory depth. Section 3 specifies the score, the quantile gate, and the view-count and angle controls. The supplementary material gives the remaining proofs, the 3×3 eigenvalue formula used in code, further ablation numbers, and held-out renders for seeds 1 and 2.

REFERENCES

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

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*, 2024.

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 *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Antoine Guédon 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 (CVPR)*, 2024.

Alex Hanson, Allen Tu, Vasu Singla, Mayuka Jayawardhana, Matthias Zwicker, and Tom Goldstein. PUP 3D-GS: Principled uncertainty pruning for 3D Gaussian splatting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Lukas Höllein, Aljaž Božič, Michael Zollhöfer, and Matthias Nießner. 3DGS-LM: Faster gaussian-splatting optimization with levenberg–marquardt. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2D Gaussian splatting for geometrically accurate radiance fields. In *ACM Transactions on Graphics (SIGGRAPH)*, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. In *ACM Transactions on Graphics (SIGGRAPH)*, 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*, 43(4), 2024.

Vincent Leroy, Yohann Cabon, and Jérôme Revaud. MASt3R: Grounding image matching in 3D with MASt3R. In *European Conference on Computer Vision (ECCV)*, 2024.

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 (ICCV)*, 2023.

Zhuoxiao Li, Shanliang Yao, Taoyu Wu, Yong Yue, Wufan Zhao, Rongjun Qin, Ángel F. García-Fernández, Andrew Levers, Jason Ralph, and Xiaohui Zhu. ULSR-GS: Urban large-scale surface reconstruction gaussian splatting with multi-view geometric consistency. *ISPRS Journal of Photogrammetry and Remote Sensing*, 230:861–880, 2025.

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 (CVPR)*, 2024.

Yang Liu, He Guan, 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 (ECCV)*, 2024.

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 (ICLR)*, 2025.

Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-GS: Structured 3D gaussians for view-adaptive rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020.

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

Linfei Pan, Daniel Barath, Marc Pollefeys, and Johannes L. Schönberger. GLOMAP: Global structure-from-motion revisited. In *European Conference on Computer Vision (ECCV)*, 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*, 2025.

Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Bill Triggs, Philip F. McLauchlan, Richard I. Hartley, and Andrew W. Fitzgibbon. Bundle adjustment—a modern synthesis. In *Vision Algorithms: Theory and Practice*, pp. 298–372. Springer, 2000.

Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. VGGT: Visual geometry grounded transformer. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jérôme Revaud. DUSt3R: Geometric 3D vision made easy. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Jianing Yang, Alexander Sax, Kevin J. Liang, Mikael Henaff, Hao Tang, Ang Cao, Joyce Chai, Franziska Meier, and Matt Feiszli. Fast3R: Towards 3D reconstruction of 1000+ images in one forward pass. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Zongxin Ye, Wenyu Li, Sidun Liu, Peng Qiao, and Yong Dou. AbsGS: Recovering fine details for 3D Gaussian splatting. In *ACM International Conference on Multimedia*, 2024.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-splatting: Alias-free 3D gaussian splatting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Hexu Zhao, Haoyang Weng, Daohan Lu, Ang Li, Jinyang Li, Aurojit Panda, and Saining Xie. On scaling up 3D gaussian splatting training. In *International Conference on Learning Representations (ICLR)*, 2025.

A Pointer to the supplementary material

Proofs of the one-view kernel, of Proposition 1, and of the gate bounds are in the supplementary material, together with the bearing-projector calculation, the eigenvalue formula, additional ablation numbers, and the seed-1 and seed-2 renders. Depth maps in the evaluator are OpenEXR values divided by 100, the MatrixCity centimetre convention. Valid depth excludes non-finite values, depths outside $(0.1, 1000)\,\mathrm{m}$, and pixels whose 3×3 morphological range exceeds $\max(1, 0.05 \times \mathrm{depth})$. Opacity pruning uses the vendor threshold 0.005. The opacity-reset interval is $3{,}000$ steps, longer than the $2{,}000$-step run, so no reset occurs.