

SkipGS: Post-Densification Backward Skipping for Efficient 3DGS Training

Jingxing Li[✉], Yongjae Lee[✉], and Deliang Fan[✉]

Arizona State University, Tempe, AZ 85281, USA
{jingxing,ylee298,dfan}@asu.edu

Abstract. 3D Gaussian Splatting (3DGS) achieves real-time novel-view synthesis by optimizing millions of anisotropic Gaussians, yet its training remains expensive, with the backward pass dominating runtime in the post-densification refinement phase. We observe substantial update redundancy in this phase: many sampled views have near-plateaued losses and provide diminishing gradient benefits, but standard training still runs full backpropagation. We propose *SkipGS* with a novel view-adaptive backward gating mechanism for efficient post-densification training. SkipGS always performs the forward pass to update per-view loss statistics, and selectively skips backward passes when the sampled view’s loss is consistent with its recent per-view baseline, while enforcing a minimum backward budget for stable optimization. On Mip-NeRF 360, compared to 3DGS, SkipGS reduces end-to-end training time by **23.1%**, driven by a **42.0%** reduction in post-densification time, with comparable reconstruction quality. Because it only changes when to backpropagate without modifying the renderer, representation, or loss, SkipGS is plug-and-play and compatible with other complementary efficiency strategies, enabling additive speedups. Code is available at <https://github.com/ASU-ESIC-FAN-Lab/SkipGS>.

Keywords: 3D Gaussian Splatting · Training acceleration · Novel-view synthesis

1 Introduction

3D Gaussian Splatting (3DGS) [5] has become the state-of-the-art (SOTA) baseline for real-time novel-view synthesis by optimizing millions of anisotropic Gaussians with a fully differentiable rasterizer. Despite its rendering efficiency, training remains expensive: modern 3DGS pipelines typically require tens of thousands of iterations, and the backward pass dominates runtime once the Gaussian set grows to a saturated number of primitives. This training cost is a practical bottleneck for interactive scene capture, large-scale benchmarking, and downstream pipelines that require repeated 3DGS fitting.

Most existing efforts to accelerate 3DGS training focus on the *primitive dimension*: they reduce the number of Gaussians via pruning/compaction, or constrain Gaussian growth during densification under a resource budget. While effective, these approaches primarily optimize *how many* primitives are processed per iteration.

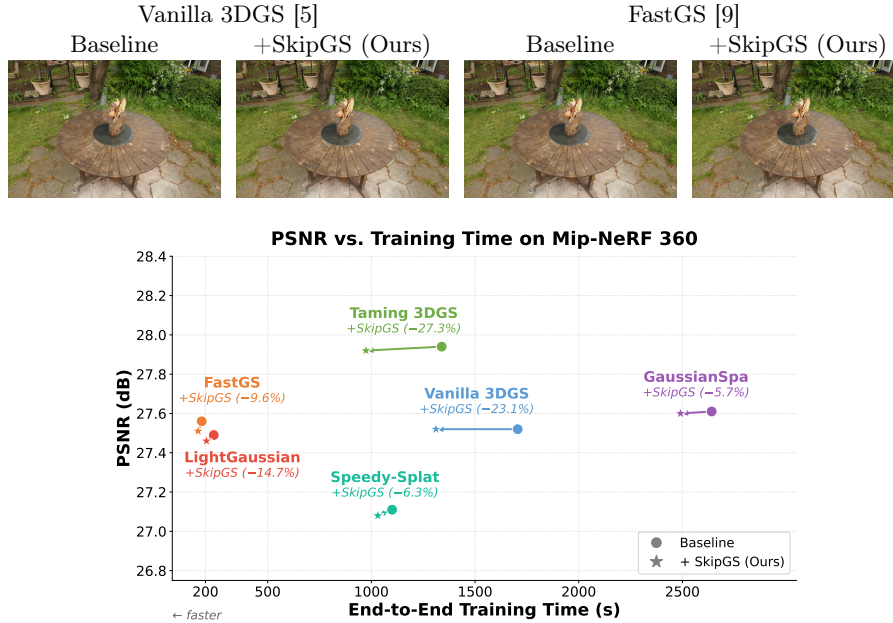


Fig. 1: SkipGS accelerates diverse 3DGS pipelines with negligible quality loss. *Top:* Qualitative comparison on the *garden* scene (Mip-NeRF 360 [1]). Baseline and +SkipGS renderings are visually indistinguishable. *Bottom:* PSNR vs. training time for all six baselines.

They leave largely unexplored an orthogonal opportunity arising from a distinctive property of 3DGS training: its two-phase structure induced by densification.

During the early *densification* phase, the Gaussian set expands rapidly and optimization must frequently revisit many views to establish geometry and appearance. In the subsequent *post-densification* phase, the Gaussian set becomes fixed and training turns into a long refinement stage. In this regime, we observe substantial update redundancy: per-view losses stabilize at different rates, and many sampled views yield diminishing returns, yet standard 3DGS [5] still executes a full backward pass whenever a view is sampled.

We quantify this redundancy through profiling analysis of vanilla 3DGS training. As shown in Fig. 2(a), the backward pass dominates per-iteration cost ($\sim 62\%$) after densification stops, making it the primary target for acceleration. Meanwhile, Fig. 2(b) reveals that per-Gaussian gradient norms decrease $\sim 2\times$ from early to late training and become nearly flat after T_d , while average Adam update norms remain comparatively stable due to momentum inertia, indicating that many post-densification backward passes produce weakly informative gradients that contribute little to parameter updates. Together, these observations suggest that a significant fraction of post-densification backward passes are re-

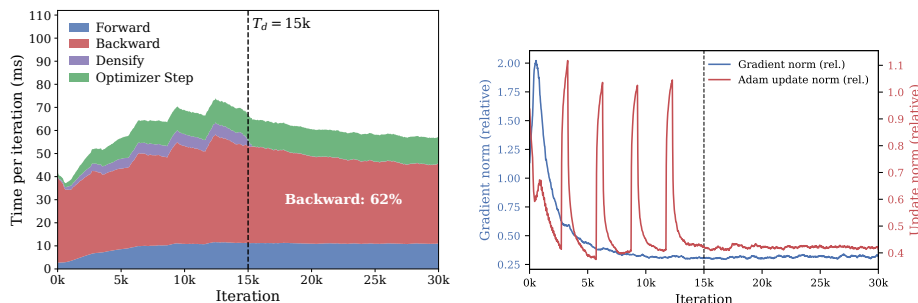


Fig. 2: Profiling vanilla 3DGS training on the *Kitchen* scene (Mip-NeRF 360). (a) Per-iteration time breakdown: the backward pass dominates ($\sim 62\%$) after densification stops at $T_d=15k$, motivating backward-level acceleration. (b) Average per-Gaussian gradient norms (blue, left axis) decrease $\sim 2\times$ from early to late training and become nearly flat after T_d , while average Adam update norms (red, right axis) remain comparatively stable (only $\sim 1.2\times$ reduction overall) due to momentum inertia, suggesting many post-densification updates are weakly informative and can be reduced by selective backpropagation. In (b), both norms are normalized by their respective values at iteration $T_d=15k$ for cross-quantity comparability.

dundant: the dominant cost component yields diminishing optimization benefit. This motivates a question: can we reduce training time by executing backward passes only when a sampled view is expected to provide useful optimization signal?

We propose *SkipGS*, a view-adaptive backward gating mechanism for post-densification 3DGS training. SkipGS retains the standard forward pass to compute the loss and update per-view statistics, but selectively skips backward passes when the sampled view’s loss is consistent with its recent per-view baseline. To ensure stable optimization, we include warmup initialization and enforce a minimum backward budget. Because our method operates at the level of *when* to backpropagate rather than changing the renderer, Gaussian representation, or loss/optimizer formulation, it is plug-and-play for existing 3DGS variants and compatible with other complementary acceleration directions such as pruning and budgeted densification [3, 8, 9], yielding additive speedups while preserving reconstruction quality.

The major contributions of this work include:

- **Post-densification backward skipping.** We propose *SkipGS*, a view-adaptive backward gating mechanism for the post-densification phase of 3DGS. SkipGS always runs the forward pass to track per-view loss statistics, and skips redundant backward passes when the sampled view’s loss is consistent with its recent per-view baseline, while enforcing stability with a minimum backward budget.
- **Speedups with preserved quality.** We show that SkipGS yields substantial wall-clock training speedups in the post-densification phase without compromising reconstruction quality. On Mip-NeRF 360, it reduces end-to-end

training time by **23.1%**, driven by a **42.0%** reduction in post-densification time over 3DGS, while maintaining comparable rendering quality.

- **Orthogonal compatibility.** Because SkipGS operates solely at the backward gating level without modifying the renderer, representation, or loss, it can be applied on top of existing 3DGS variants as a post-densification plug-in. We empirically demonstrate that SkipGS *further accelerates* representative state-of-the-art efficient 3DGS methods, including FastGS [9], Taming 3DGS [8], GaussianSpa [14], LightGaussian [2], and Speedy-Splat [3], yielding additional wall-clock savings while maintaining comparable reconstruction quality across three standard benchmarks.

2 Related Work

3D Gaussian Splatting. 3D Gaussian Splatting (3DGS) [5] represents scenes as anisotropic Gaussians and enables efficient differentiable rendering with a tile-based rasterizer. Its training pipeline naturally follows a two-phase structure: a densification phase that grows and refines the Gaussian set, followed by a post-densification refinement phase where the Gaussian set is fixed and parameters are fine-tuned. We leverage this phase transition to reduce redundant computation without changing the representation.

Gaussian compaction. A dominant direction improves efficiency by explicitly reducing the number of Gaussians, which lowers rendering cost and can reduce the per-iteration workload in optimization. FastGS [9] integrates training-time pruning with pipeline optimizations and explicitly targets wall-clock training acceleration. GaussianSpa [14] progressively enforces sparsity (e.g., via opacity- or importance-based criteria) during training to obtain a compact Gaussian set. LightGaussian [2] performs post-training pruning and subsequent fine-tuning to compress the Gaussian set while maintaining reconstruction quality. Speedy-Splat [3] optimizes the rendering pipeline for precise Gaussian localization and integrates a training-time pruning technique, jointly improving rendering speed, model compactness, and training efficiency. Overall, these methods primarily optimize *how many* Gaussians participate in rendering and learning, often improving model compactness and rendering efficiency; the impact on end-to-end training time depends on the specific pipeline and protocol.

Gaussian growth control. Another line of work targets the *growth dynamics* of densification by regulating the Gaussian count under a resource budget, rather than removing Gaussians after the fact. Taming 3DGS [8] studies training under compute/memory budgets and proposes mechanisms that control densification and optimization under constrained resources. These methods operate at the representation-growth level and primarily optimize *how large* the Gaussian set becomes (and thus the per-iteration workload), with some works explicitly reporting training-time speedups.

Our position: backward gating from the two-phase training structure. Existing efficiency improvements largely optimize the *primitive dimension* of 3DGS training by either reducing the Gaussian set or controlling its growth during densification. In contrast, we target a largely overlooked axis: *whether* to execute the backward pass at each iteration. SkipGS exploits per-view convergence heterogeneity in post-densification training and gates redundant backward passes based on per-view loss deviation, while enforcing stability budgets.

3 Preliminaries: 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [5] represents a scene as a set of N learnable anisotropic 3D Gaussians and renders novel views through a fully differentiable, tile-based rasterizer. Each Gaussian G_i is defined by a center $\boldsymbol{\mu}_i \in \mathbb{R}^3$, a 3D covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{3 \times 3}$, an opacity $\alpha_i \in [0, 1]$, and a set of spherical-harmonic (SH) coefficients $\mathbf{f}_i$ encoding view-dependent color. The Gaussian density at a point $\mathbf{x}$ is

$$G_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right). \quad (1)$$

To guarantee positive semi-definiteness during gradient-based optimization, $\boldsymbol{\Sigma}_i$ is factorized as

$$\boldsymbol{\Sigma}_i = \mathbf{R}_i \mathbf{S}_i \mathbf{S}_i^\top \mathbf{R}_i^\top, \quad (2)$$

where $\mathbf{R}_i$ is a rotation matrix stored as a unit quaternion $\mathbf{q}_i$ and $\mathbf{S}_i = \text{diag}(s_{i,1}, s_{i,2}, s_{i,3})$ is a diagonal scaling matrix whose entries represent the per-axis standard deviations of the Gaussian.

Given a camera with viewing transform $\mathbf{W}$ and the Jacobian $\mathbf{J}$ of the local affine approximation to the projective mapping [15], the 2D screen-space covariance is

$$\boldsymbol{\Sigma}_i^{2D} = \mathbf{J} \mathbf{W} \boldsymbol{\Sigma}_i \mathbf{W}^\top \mathbf{J}^\top. \quad (3)$$

Pixel colors are composited by front-to-back alpha blending over the depth-sorted Gaussians overlapping a given pixel $\mathbf{p}$:

$$C(\mathbf{p}) = \sum_{i \in \mathcal{N}} c_i \alpha'_i T_i, \quad T_i = \prod_{j=1}^{i-1} (1 - \alpha'_j), \quad (4)$$

where c_i is the SH-evaluated color for the current view direction, T_i denotes the transmittance accumulated from all Gaussians in front of G_i , and the effective per-pixel opacity is

$$\alpha'_i = \alpha_i \exp\left(-\frac{1}{2} \Delta \mathbf{p}_i^\top (\boldsymbol{\Sigma}_i^{2D})^{-1} \Delta \mathbf{p}_i\right), \quad \Delta \mathbf{p}_i = \mathbf{p} - \boldsymbol{\mu}_i^{2D}. \quad (5)$$

The Gaussian parameters are initialized from a Structure-from-Motion [10] point cloud and optimized with Adam [6] under a photometric loss

$$\mathcal{L} = (1 - \lambda) \mathcal{L}_1 + \lambda (1 - \text{SSIM}), \quad (6)$$

where $\lambda = 0.2$ and SSIM is the structural similarity index [11]. During training, an adaptive density control scheme periodically *clones* under-reconstructed Gaussians (high view-space positional gradient, small spatial extent), *splits* over-reconstructed ones (high gradient, large extent), and *prunes* near-transparent Gaussians ($\alpha_i < \epsilon_\alpha$, where ϵ_α is a small opacity threshold).

Crucially, 3DGS [5] training follows a two-phase training process induced by densification:

- **Densification phase** (iterations $0-T_d$, $T_d=15k$ by default): densification is active and the Gaussian count grows rapidly.
- **Post-densification phase** (iterations T_d-T , $T=30k$ by default): densification is disabled; the Gaussian set is *fixed* (typically $N = 1M-5M$) and parameters are refined.

4 Proposed Method

4.1 Overview: Constrained Backward Gating

Within this section, t denotes the iteration index relative to the start of the post-densification phase (i.e., $t = 1$ at the first iteration after T_d). We focus on the post-densification phase. In this phase, per-view losses often stabilize at different rates: a subset of views continues to provide useful gradients while many others are repeatedly revisited with little progress. Yet standard 3DGS [5] still performs a full backward pass whenever a view is sampled. Since backward dominates the per-iteration cost, this update redundancy in the late-phase leads to substantial wasted computation. This observation motivates our backward gating mechanism, which decides for each sampled view whether to skip or execute the backward pass based on its recent loss deviation.

SkipGS targets **post-densification** training, where the Gaussian set is fixed and optimization mainly fine-tunes appearance and geometry. Although many training views have already converged in this regime, standard 3DGS [5] still performs a full backward pass on every sampled view, even when the view’s loss is stable relative to its recent per-view baseline. We formulate post-densification acceleration as a **constrained backward gating** problem. At iteration t , a view v_t is sampled and we always perform the forward pass to compute $\mathcal{L}_{v_t}^{(t)}$ and update per-view statistics. We then decide whether to execute the backward pass so as to (i) focus compute on views with remaining optimization potential while (ii) maintaining stable convergence by enforcing a minimum update budget.

As illustrated in Fig. 3, SkipGS implements this gating mechanism with two ingredients:

1. **Per-view deviation score and skip test** (Sec. 4.2): For each sampled view, we compute a scale-invariant deviation score that compares its current loss against its recent EMA baseline. Backward is executed only when the score exceeds 1, indicating the view’s loss has risen above its baseline and is likely to benefit from gradient computation.

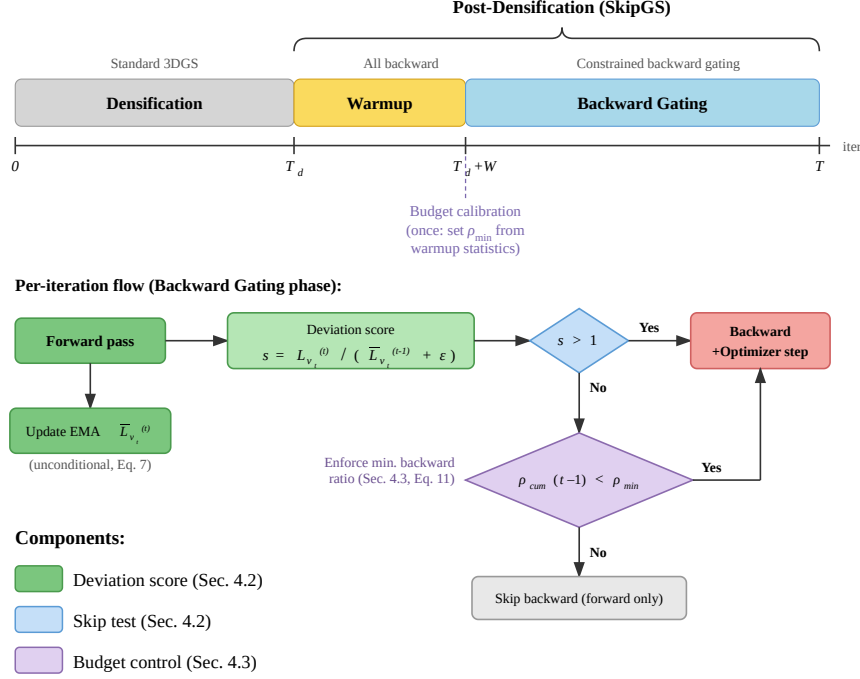


Fig. 3: Overview of SkipGS. *Top:* Training timeline. SkipGS activates after densification stops at T_d : a warmup window (W iterations) initializes a per-view exponential moving average (EMA) baseline and calibrates the minimum backward budget $\rho_{\min}$, after which constrained backward gating begins. *Bottom:* Per-iteration decision flow during the backward gating phase. At each iteration, the forward pass is always executed to compute the loss and update the per-view EMA $\bar{\mathcal{L}}_{v_t}^{(t)}$ (unconditional). A **deviation score** $s = \mathcal{L}_{v_t}^{(t)} / (\bar{\mathcal{L}}_{v_t}^{(t-1)} + \epsilon)$ measures whether the current loss exceeds the recent per-view baseline (Sec. 4.2). If $s > 1$, the backward pass is executed; otherwise, SkipGS proposes to skip. Before skipping, the **budget controller** checks whether the cumulative backward ratio ρ_{cum} has fallen below the auto-calibrated minimum $\rho_{\min}$, and forces backward execution if so (Sec. 4.3). Only when both checks allow skipping is the backward pass omitted.

- 2. Backward budget calibration and control (Sec. 4.3):** To prevent gradient starvation from over-aggressive skipping, we enforce a minimum backward ratio $\rho_{\min}$, auto-calibrated from warmup statistics. When the cumulative backward ratio ρ_{cum} falls below $\rho_{\min}$, the skip decision is overridden and backward is forced.

Algorithm 1 summarizes the overall procedure. Last but not least, SkipGS only changes *when* to backpropagate; it does not modify the renderer, the Gaussian representation, or the loss. It is therefore orthogonal to existing accelerations

Algorithm 1 SkipGS: Post-Densification Backward Gating

Require: Total iterations T , densification stop iteration T_d , warmup length W , EMA decay β , stability constant ϵ , budget floor ρ_{lo}

- 1: $T_{\text{post}} \leftarrow T - T_d$
- 2: Initialize per-view EMA $\bar{\mathcal{L}}[\cdot] \leftarrow \emptyset$; $\rho_{\min} \leftarrow \rho_{lo}$; backward count $b \leftarrow 0$
- 3: **for** $t = 1$ **to** T_{post} **do**
- 4: Sample view v_t ; compute forward loss $\mathcal{L}_{v_t}^{(t)}$
- 5: **Deviation score:** $s \leftarrow \mathcal{L}_{v_t}^{(t)} / (\bar{\mathcal{L}}_{v_t}^{(t-1)} + \epsilon)$ if $\bar{\mathcal{L}}_{v_t}^{(t-1)}$ exists, else $s \leftarrow +\infty$
- 6: **Skip test:** $g \leftarrow \mathbf{1}[s > 1]$
- 7: $\rho_{\text{cum}}(t-1) \leftarrow b / \max(t-1, 1)$ $\triangleright$ cumulative backward ratio before iteration t
- 8: **if** $t \leq W$ **then** $\triangleright$ warmup: always backward, collect stats
- 9: Record g for budget calibration; force $g \leftarrow 1$
- 10: **else**
- 11: Calibrate $\rho_{\min}$ from warmup stats (once, at $t=W+1$; Eq. 13)
- 12: **If** $\rho_{\text{cum}}(t-1) < \rho_{\min}$, force $g \leftarrow 1$ $\triangleright$ budget override
- 13: **end if**
- 14: **if** $g = 1$ **then**
- 15: Backward; optimizer step; $b \leftarrow b + 1$
- 16: **end if**
- 17: Update $\bar{\mathcal{L}}_{v_t}^{(t)} \leftarrow \beta \bar{\mathcal{L}}_{v_t}^{(t-1)} + (1-\beta) \mathcal{L}_{v_t}^{(t)}$ $\triangleright$ unconditional (Eq. 7)
- 18: **end for**

such as Gaussian compaction and methods that regulate Gaussian count growth during densification, and can be combined with them for additive speedups while maintaining rendering quality.

4.2 Per-View Deviation Score

After densification is disabled, different views converge at different rates due to varying visibility, occlusion patterns, and view-dependent appearance. A single global loss trend is therefore insufficient to identify which sampled views still provide useful optimization signal. We construct a lightweight per-view deviation score from recent loss statistics.

For each training view, we maintain an exponential moving average (EMA) of its observed loss. When view v_t is sampled at iteration t ,

$$\bar{\mathcal{L}}_{v_t}^{(t)} = \beta \bar{\mathcal{L}}_{v_t}^{(t-1)} + (1 - \beta) \mathcal{L}_{v_t}^{(t)}, \quad (7)$$

with $\beta \in (0, 1)$; unsampled views keep their previous EMA. We then measure the normalized deviation of v_t from this baseline by a scale-invariant ratio:

$$s_{v_t}^{(t)} = \frac{\mathcal{L}_{v_t}^{(t)}}{\bar{\mathcal{L}}_{v_t}^{(t-1)} + \epsilon}, \quad (8)$$

where $\epsilon = 10^{-8}$ is a small constant for numerical stability. In implementation, we update the per-view EMA after each forward observation and deviation score

calculation, regardless of whether the backward pass is executed. Thus, $\mathcal{L}_{v_t}^{(t)}$ always denotes the observed forward loss, ensuring the deviation score tracks per-view loss trends independent of gradient updates.

We refer to $s_{v_t}^{(t)}$ as the *deviation score*. Values significantly above 1 indicate a loss spike relative to the recent baseline, suggesting that the view remains under-optimized (or has regressed) and is likely to benefit from backward computation. Values below 1 indicate the view currently performs better than its recent baseline; in this case, backward passes typically yield diminishing returns, and we prioritize compute for high-deviation views.

Given the deviation score, we decide whether to execute backward via a skip test:

$$g_t(v_t) = \mathbf{1}\left[s_{v_t}^{(t)} > 1\right], \quad (9)$$

where $g_t(v_t) = 1$ triggers a backward pass and $g_t(v_t) = 0$ skips it. We backpropagate only when the view’s current loss exceeds its recent EMA baseline, i.e., when the deviation score is above 1.

To allow per-view EMAs to stabilize after densification ends, we apply a warmup window of W iterations during which all backward passes are unconditionally executed. During warmup, we still evaluate the skip test for budget calibration (Sec. 4.3) and collect statistics only when the sampled view already has an EMA history.

4.3 Backward Budget Calibration and Control

The view-adaptive backward gating in Eq. (9) proposes skipping on low-utility views, but unconstrained skipping can starve optimization of gradient signal. We therefore enforce a **minimum backward budget** on average: when the cumulative backward ratio since the start of the post-densification phase falls below a target minimum, we override the gating decision and force backward execution.

Let $g_t \in \{0, 1\}$ indicate whether post-densification iteration t executes backward. We maintain the cumulative backward ratio

$$\rho_{\text{cum}}(t) = \frac{1}{t} \sum_{i=1}^t g_i. \quad (10)$$

Since g_t is decided online, we apply the budget-floor check using the ratio *before* making the current decision:

$$\text{if } \rho_{\text{cum}}(t-1) < \rho_{\min} \text{ then force } g_t \leftarrow 1. \quad (11)$$

Because warmup iterations execute backward ($g_t=1$), including warmup in ρ_{cum} yields a smooth budget signal that is less sensitive to short-term fluctuations.

The appropriate minimum budget depends on scene difficulty and densification quality. To avoid per-scene tuning, we calibrate $\rho_{\min}$ from warmup statistics.

During warmup (backward always executed), we still evaluate the gating criterion in Eq. (9) and record the fraction of iterations that would trigger backward:

$$\hat{\rho}_W = \frac{1}{|\mathcal{T}_W|} \sum_{i \in \mathcal{T}_W} g_i(v_i), \quad (12)$$

where $\mathcal{T}_W$ denotes warmup iterations whose sampled view already has an EMA history. Equivalently, $1 - \hat{\rho}_W$ gives the fraction of warmup iterations where the gating would have proposed skipping ($s \leq 1$).

We then set the minimum backward ratio by a lower-bounded linear interpolation:

$$\rho_{\min} = \rho_{\text{lo}} + (1 - \rho_{\text{lo}}) \hat{\rho}_W, \quad (13)$$

where $\rho_{\text{lo}} = 0.5$ in all experiments. Intuitively, if warmup indicates many views are already easy (small $\hat{\rho}_W$), we allow a lower minimum budget; if warmup suggests most views still require backward ($\hat{\rho}_W \approx 1$), we enforce a budget close to full backward.

5 Experiments

We evaluate *SkipGS* on standard 3DGS [5] benchmarks to assess its ability to reduce wall-clock training time in the post-densification phase while preserving reconstruction quality. We report results on the vanilla 3DGS pipeline and on representative acceleration baselines that either (i) reduce the number of Gaussians (pruning/compaction) or (ii) control Gaussian growth during densification. For each baseline, we additionally report the performance of *baseline + SkipGS (ours)* to quantify the additive benefit and verify composability.

5.1 Experimental Setup

Datasets. Following the standard 3DGS [5] protocol, we conduct experiments on three real-world datasets: Mip-NeRF 360 [1], Deep Blending [4], and Tanks and Temples [7]. Unless otherwise specified, we train each scene for 30K iterations with densification enabled until $T_d = 15\text{K}$, followed by post-densification fine-tuning.

Baselines and evaluation protocol. We compare against vanilla 3DGS [5] and representative efficiency-oriented baselines from two families: (i) *Gaussian compaction* methods that reduce the number of Gaussians (e.g., FastGS [9], GaussianSpa [14], LightGaussian [2], Speedy-Splat [3]); and (ii) *Gaussian growth control* methods that regulate densification under constraints (e.g., Taming 3DGS [8]). For each baseline, we evaluate both the original method and the combined setting *baseline + SkipGS*, where our method is enabled only in the post-densification phase and all other components of the baseline remain unchanged.

Metrics. We report standard novel-view synthesis metrics, including PSNR, SSIM [11], and LPIPS [13]. Training efficiency is evaluated by the end-to-end wall-clock training time (T_{total}) as well as the post-densification refinement time (T_{post}), both measured in seconds.

Implementation details. All experiments are run on a single NVIDIA RTX Ada 5000 (32 GB) GPU. Unless a baseline requires otherwise, we use the standard 3DGS [5] optimizer (Adam) and loss as in [5]. SkipGS is enabled only after densification stops (post-densification phase): we always execute the forward pass, maintain per-view loss EMAs, and decide whether to execute backward via the view-adaptive backward gating (Eq. (9)). We use the same hyperparameters across all scenes and datasets in our evaluation: warmup length $W=500$, EMA decay $\beta=0.95$, stability constant $\epsilon=10^{-8}$, and ρ_{min} auto-calibrated at the end of warmup from the observed natural convergence rate of each scene. When backward is executed, we apply the standard optimizer update. We report end-to-end wall-clock speedups for the full 30K-iteration training unless stated otherwise.

5.2 Main Results: Training Acceleration in Post-Densification

We first evaluate SkipGS on the vanilla 3DGS pipeline [5]. Across all datasets, SkipGS reduces end-to-end wall-clock training time by skipping redundant backward passes in the post-densification phase, while preserving reconstruction quality measured by PSNR/SSIM/LPIPS. On Mip-NeRF 360 (Table 1), compared to vanilla 3DGS, SkipGS reduces the end-to-end training time by 23.1%, driven by a 42.0% reduction in post-densification time T_{post} . Please note that we report $T_{\text{total}} = T_{\text{post}}$ for the LightGaussian rows because LightGaussian performs prune+finetune after completing a full 3DGS training run. Notably, the final number of Gaussians is unchanged (2.739M on average), confirming that the gains come from backward gating rather than primitive reduction.

Tables 2 and 3 report the same comparison on Deep Blending and Tanks&Temples. We observe consistent reductions in T_{post} , which translate to end-to-end speedups without changing the Gaussian set.

Notably, FastGS + SkipGS achieves the fastest post-densification training across all three benchmarks (e.g., $T_{\text{post}}=69.8\text{ s}$ on Mip-NeRF 360), demonstrating that SkipGS can further accelerate the current fastest 3DGS training pipeline.

Beyond vanilla 3DGS, we apply SkipGS as a post-densification plug-in on top of representative baselines that (i) reduce or compact the Gaussian set (e.g., pruning/sparsification/compaction) or (ii) regulate Gaussian growth during densification under resource constraints. Tables 1–3 report each baseline and its *baseline + SkipGS* counterpart. Across these baselines, SkipGS provides additional wall-clock savings by targeting an orthogonal axis: it reduces *how often* backward is executed in post-densification training, whereas prior methods mainly affect *how many* primitives are processed per iteration and/or memory/throughput constraints. Importantly, SkipGS does not modify the renderer, representation, or loss, and thus can be enabled without re-tuning baseline-

Table 1: Mip-NeRF 360 [1] (avg over scenes): comparisons with 3DGS baselines and *baseline + SkipGS*. T_{total} is the end-to-end wall-clock training time (s) under each method’s protocol. T_{post} denotes the post-densification refinement time (s). LightGaussian performs prune+finetune after completing a full 3DGS training run; therefore we report only this additional post-training, so $T_{\text{total}} = T_{\text{post}}$ for LightGaussian rows. Note that SkipGS does not modify the renderer or the Gaussian set; hence the number of Gaussians as well as the rendering-time workload are identical to the corresponding baseline, and our speedups come solely from reducing post-densification backpropagation.

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$T_{\text{total}} \downarrow$	$T_{\text{post}} \downarrow$
Vanilla 3DGS	27.52	0.816	0.215	1705.7	939.6
+ SkipGS (ours)	27.52	0.816	0.217	1311.1	545.0
Δ (Ours–Vanilla)	+0.00	0.000	+0.002	-394.6 (-23.1%)	-394.6 (-42.0%)
FastGS [9]	27.56	0.798	0.261	181.9	87.2
+ SkipGS (ours)	27.51	0.797	0.262	164.5	69.8
Δ (Ours–FastGS)	-0.05	-0.001	+0.001	-17.4 (-9.6%)	-17.4 (-20.0%)
Taming 3DGS [8]	27.94	0.822	0.207	1339.0	757.0
+ SkipGS (ours)	27.92	0.822	0.209	974.0	392.0
Δ (Ours–Taming 3DGS)	-0.02	0.000	+0.002	-365.0 (-27.3%)	-365.0 (-48.2%)
GaussianSpa [14]	27.61	0.826	0.213	2640.9	1485.0
+ SkipGS (ours)	27.60	0.825	0.215	2490.9	1335.0
Δ (Ours–GaussianSpa)	-0.01	-0.001	+0.002	-150.0 (-5.7%)	-150.0 (-10.1%)
LightGaussian [2]	27.49	0.810	0.230	240.0	240.0
+ SkipGS (ours)	27.46	0.809	0.231	204.8	204.8
Δ (Ours–LightGaussian)	-0.03	-0.001	+0.001	-35.2 (-14.7%)	-35.2 (-14.7%)
Speedy-Splat [3]	27.11	0.799	0.263	1099.8	492.0
+ SkipGS (ours)	27.08	0.799	0.264	1030.8	423.0
Δ (Ours–Speedy-Splat)	-0.03	0.000	+0.001	-69.0 (-6.3%)	-69.0 (-14.0%)

specific knobs; the combined setting retains comparable reconstruction quality while improving training speed.

Why post-densification backward scheduling yields wall-clock gains. Fig. 2(a) shows that the backward pass accounts for the majority of per-iteration time throughout the post-densification phase. Meanwhile, Fig. 2(b) shows that average per-Gaussian gradient norms flatten after T_d while average Adam update norms remain stable due to momentum inertia, indicating that many post-densification backward passes produce weakly informative gradients. SkipGS exploits this post-densification redundancy by always performing the forward pass to update per-view statistics, but executing backward selectively under a minimum-budget constraint, thereby reducing the dominant post-densification cost in practice.

Qualitative results. Fig. 4 shows rendered views across three datasets and representative baselines. Despite substantial post-densification time reductions, SkipGS produces visually indistinguishable results from the corresponding full-training baseline, consistent with Tables 1–3.

Table 2: Deep Blending [4] (avg over scenes): same protocol and definitions as Table 1. For this dataset, we compute T_{total} for our method as the sum of the baseline method’s densification-phase wall-clock time and our T_{post} .

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$T_{\text{total}} \downarrow$	$T_{\text{post}} \downarrow$
Vanilla 3DGS	29.79	0.907	0.238	1713.0	965.5
+ SkipGS (ours)	29.88	0.909	0.237	1349.5	602.0
Δ (Ours–Vanilla)	+0.09	+0.002	-0.001	-363.5 (-21.2%)	-363.5 (-37.6%)
FastGS [9]	30.02	0.905	0.267	111.0	49.6
+ SkipGS (ours)	30.03	0.906	0.268	100.6	39.2
Δ (Ours–FastGS)	+0.01	+0.001	+0.001	-10.4 (-9.4%)	-10.4 (-21.0%)
Taming 3DGS [8]	29.83	0.907	0.237	1110.0	620.0
+ SkipGS (ours)	29.95	0.909	0.237	804.0	314.0
Δ (Ours–Taming 3DGS)	+0.12	+0.002	0.000	-306.0 (-27.6%)	-306.0 (-49.4%)
GaussianSpa [14]	30.19	0.913	0.238	2373.5	1309.5
+ SkipGS (ours)	30.27	0.916	0.236	2251.5	1187.5
Δ (Ours–GaussianSpa)	+0.08	+0.003	-0.002	-122.0 (-5.1%)	-122.0 (-9.3%)
LightGaussian [2]	29.92	0.905	0.251	240.0	240.0
+ SkipGS (ours)	29.92	0.906	0.252	204.5	204.5
Δ (Ours–LightGaussian)	0.00	+0.001	+0.001	-35.5 (-14.8%)	-35.5 (-14.8%)
Speedy-Splat [3]	29.60	0.905	0.260	977.5	441.5
+ SkipGS (ours)	29.59	0.905	0.261	916.0	380.0
Δ (Ours–Speedy-Splat)	-0.01	0.000	+0.001	-61.5 (-6.3%)	-61.5 (-13.9%)

Table 3: Tanks&Temples [7] (avg over scenes): same protocol and definitions as Table 1. For this dataset, we compute T_{total} for our method as the sum of the baseline method’s densification-phase wall-clock time and our T_{post} .

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$T_{\text{total}} \downarrow$	$T_{\text{post}} \downarrow$
Vanilla 3DGS	23.83	0.853	0.169	993.0	545.5
+ SkipGS (ours)	23.78	0.853	0.173	792.0	344.5
Δ (Ours–Vanilla)	-0.05	0.000	+0.004	-201.0 (-20.2%)	-201.0 (-36.8%)
FastGS [9]	24.25	0.843	0.208	106.0	46.1
+ SkipGS (ours)	24.17	0.842	0.210	98.3	38.4
Δ (Ours–FastGS)	-0.08	-0.001	+0.002	-7.7 (-7.3%)	-7.7 (-16.7%)
Taming 3DGS [8]	24.43	0.859	0.164	774.0	424.0
+ SkipGS (ours)	24.34	0.858	0.169	596.0	246.0
Δ (Ours–Taming 3DGS)	-0.09	-0.001	+0.005	-178.0 (-23.0%)	-178.0 (-42.0%)
GaussianSpa [14]	23.71	0.854	0.168	1675.5	731.5
+ SkipGS (ours)	23.66	0.854	0.169	1600.5	656.5
Δ (Ours–GaussianSpa)	-0.05	0.000	+0.001	-75.0 (-4.5%)	-75.0 (-10.3%)
LightGaussian [2]	23.89	0.847	0.187	160.0	160.0
+ SkipGS (ours)	23.85	0.846	0.189	140.5	140.5
Δ (Ours–LightGaussian)	-0.04	-0.001	+0.002	-19.5 (-12.2%)	-19.5 (-12.2%)
Speedy-Splat [3]	23.66	0.830	0.222	600.5	267.5
+ SkipGS (ours)	23.60	0.829	0.223	568.2	235.2
Δ (Ours–Speedy-Splat)	-0.06	-0.001	+0.001	-32.3 (-5.4%)	-32.3 (-12.1%)

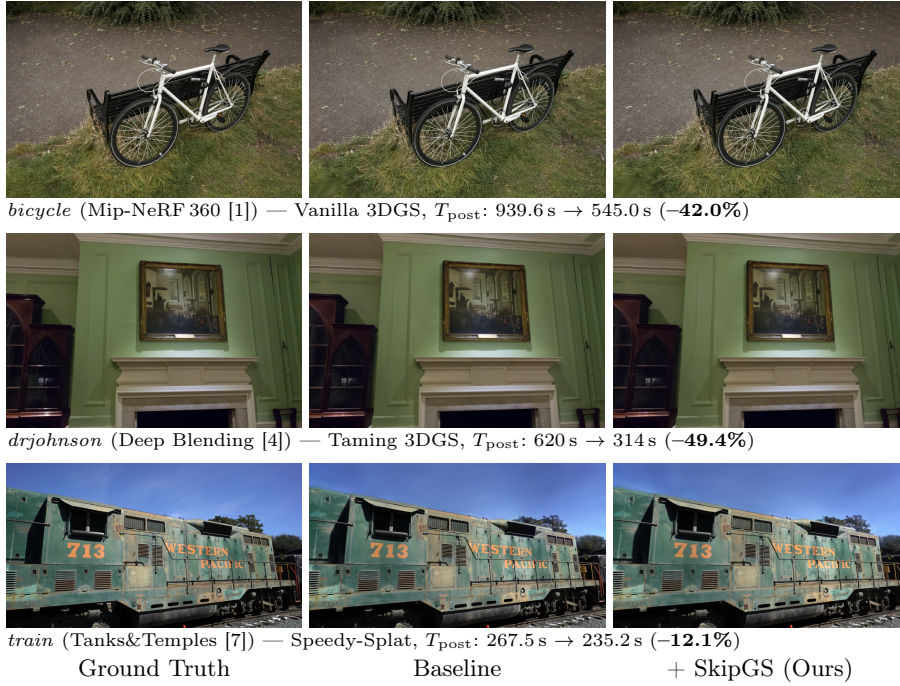


Fig. 4: Qualitative comparison across datasets and baselines. Each row shows a different dataset and baseline method. Despite substantial reductions in post-densification time, SkipGS produces visually indistinguishable results from the corresponding full-training baseline across all settings.

5.3 Ablation Study

We ablate the backward budget control mechanism (Sec. 4.3) on Mip-NeRF 360 [1] using two representative baselines: GaussianSpa [14] and Taming 3DGS [8]. Table 4 compares three configurations for each: the baseline without SkipGS, the full SkipGS pipeline, and SkipGS with budget control removed.

Effect of backward budget control. Across both baselines, the full SkipGS pipeline achieves substantial post-densification speedups while preserving quality: PSNR drops by only 0.01 dB on GaussianSpa and 0.02 dB on Taming 3DGS, with SSIM and LPIPS nearly unchanged. Removing budget control yields further time savings (e.g., T_{post} : 757 s $\rightarrow$ 161 s for Taming 3DGS), but at severe quality cost: PSNR drops by 0.67 dB and 0.38 dB on Taming 3DGS and GaussianSpa, respectively, with large degradation in SSIM and LPIPS. Without the budget mechanism, the gating becomes overly aggressive; too many backward passes are skipped, starving the optimizer of gradient signal. The budget controller prevents this by forcing backward execution when the cumulative ratio falls below $\rho_{\min}$ (Eq. 13), recovering nearly all quality while retaining the majority of the

Table 4: Ablation on Mip-NeRF 360 [1] (avg over scenes). Removing budget control leads to over-aggressive skipping: T_{post} drops further but quality degrades substantially. The budget mechanism trades a modest time increase for quality preservation.

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$T_{\text{post}}(\text{s})\downarrow$
GaussianSpa (baseline)	27.61	0.826	0.213	1485.0
+ SkipGS (full)	27.60	0.825	0.215	1335.0
Δ (full-baseline)	-0.01	-0.001	+0.002	-150.0
+ SkipGS w/o budget	27.23	0.804	0.249	1026.0
Δ (w/o-baseline)	-0.38	-0.022	+0.036	-459.0
Taming 3DGS (baseline)	27.94	0.822	0.207	757.0
+ SkipGS (full)	27.92	0.822	0.209	392.0
Δ (full-baseline)	-0.02	0.000	+0.002	-365.0
+ SkipGS w/o budget	27.27	0.795	0.260	161.0
Δ (w/o-baseline)	-0.67	-0.027	+0.053	-596.0

speedup. This confirms that budget control is essential for balancing acceleration and quality preservation across different base methods.

6 Conclusion

We presented *SkipGS*, a view-adaptive backward gating mechanism for the post-densification phase of 3D Gaussian Splatting [5]. By tracking per-view loss statistics and selectively skipping backward passes when a view’s loss is consistent with its recent baseline, SkipGS reduces redundant computation where the backward pass dominates runtime. A warmup-calibrated minimum backward budget preserves stable optimization. Experiments on Mip-NeRF 360 [1], Deep Blending [4], and Tanks&Temples [7] show consistent wall-clock reductions with comparable reconstruction quality. Last but not least, our experiments are limited to static scenes, and we leave the evaluation of SkipGS on dynamic scenes and models [12] for future work.

Acknowledgment

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/ Interior Business Center (DOI/IBC) contract number 140D0423C0076. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI/IBC, or the U.S. Government.

References

1. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5470–5479 (June 2022)
2. Fan, Z., Wang, K., Wen, K., Zhu, Z., Xu, D., Wang, Z.: Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. In: Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C. (eds.) *Advances in Neural Information Processing Systems*. vol. 37, pp. 140138–140158. Curran Associates, Inc. (2024). <https://doi.org/10.52202/079017-4447>, https://proceedings.neurips.cc/paper_files/paper/2024/file/fd881d3b625437354d4421818f81058f-Paper-Conference.pdf
3. Hanson, A., Tu, A., Lin, G., Singla, V., Zwicker, M., Goldstein, T.: Speedy-splat: Fast 3d gaussian splatting with sparse pixels and sparse primitives. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 21537–21546 (June 2025)
4. Hedman, P., Philip, J., Price, T., Frahm, J.M., Drettakis, G., Brostow, G.: Deep blending for free-viewpoint image-based rendering. *ACM Trans. Graph.* **37**(6) (Dec 2018). <https://doi.org/10.1145/3272127.3275084>, <https://doi.org/10.1145/3272127.3275084>
5. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* **42**(4), 139:1–139:14 (2023). <https://doi.org/10.1145/3592433>, <https://doi.org/10.1145/3592433>
6. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015), <http://arxiv.org/abs/1412.6980>
7. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and temples: benchmarking large-scale scene reconstruction. *ACM Trans. Graph.* **36**(4) (Jul 2017). <https://doi.org/10.1145/3072959.3073599>, <https://doi.org/10.1145/3072959.3073599>
8. Mallick, S.S., Goel, R., Kerbl, B., Steinberger, M., Carrasco, F.V., De La Torre, F.: Taming 3dgs: High-quality radiance fields with limited resources. In: *SIGGRAPH Asia 2024 Conference Papers*. SA '24, Association for Computing Machinery, New York, NY, USA (2024). <https://doi.org/10.1145/3680528.3687694>, <https://doi.org/10.1145/3680528.3687694>
9. Ren, S., Wen, T., Fang, Y., Lu, B.: Fastgs: Training 3d gaussian splatting in 100 seconds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 26094–26103 (June 2026)
10. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4104–4113 (2016)
11. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004). <https://doi.org/10.1109/TIP.2003.819861>
12. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4d gaussian splatting for real-time dynamic scene rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20310–20320 (June 2024)

13. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
14. Zhang, Y., Jia, W., Niu, W., Yin, M.: Gaussianspa: An "optimizing-sparsifying" simplification framework for compact and high-quality 3d gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 26673–26682 (June 2025)
15. Zwicker, M., Pfister, H., van Baar, J., Gross, M.: Ewa splatting. *IEEE Transactions on Visualization and Computer Graphics* **8**(3), 223–238 (2002). <https://doi.org/10.1109/TVCG.2002.1021576>