

PSCD-GS: Role-Aware Perception-Structure Collaborative Densification for 3D Gaussian Splatting

Deyong Shang

shangdeyong@cumtb.edu.cn

China University of Mining and Technology - Beijing

Jiaxin Shi

China University of Mining and Technology - Beijing

Lianyu Mu

China University of Mining and Technology - Beijing

Research Article

Keywords: 3D Gaussian Splatting, novel view synthesis, Gaussian densification, density control, perceptual guidance, structural guidance

Posted Date: August 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10466466/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

PSCD-GS: Role-Aware Perception–Structure Collaborative Densification for 3D Gaussian Splatting

Deyong Shang^{1*}, Jiaxin Shi¹ and Lianyu Mu¹

^{1*}School of Mechanical and Electrical Engineering, China University of Mining and Technology-Beijing, Ding No. 11, Xueyuan Road, Haidian District, Beijing, 100083, Beijing, China.

*Corresponding author(s). E-mail(s): shangdeyong@cumtb.edu.cn;
Contributing authors: SQT2400401024@student.cumtb.edu.cn;
SQT2500401015@student.cumtb.edu.cn;

Abstract

3D Gaussian Splatting (3DGS) enables efficient, high-quality novel-view synthesis, but its reconstruction quality remains sensitive to how Gaussian primitives are densified during optimization. Existing densification strategies commonly rely on gradients, reconstruction errors, or isolated regional cues, which may not capture the complementary perceptual and structural characteristics of a scene. We propose PSCD-GS, a role-aware perception–structure collaborative densification framework that integrates image-space perceptual responses with multi-view structural evidence. These cues are projected onto individual Gaussian primitives through rendering-contribution-weighted aggregation, producing continuous memberships for perception-dominant, structure-dominant, co-dominant, and low-response behaviors. This primitive-level representation is used to prioritize densification candidates, select between cloning and splitting, and guide the placement of newly generated Gaussians. Experiments on Mip-NeRF 360, Deep Blending, and Tanks & Temples show that PSCD-GS achieves competitive dataset-level rendering quality while maintaining a controlled Gaussian representation. Qualitative comparisons suggest better preservation of fine textures, object boundaries, and thin structures. These results indicate that jointly modeling perceptual and structural evidence can provide more informative guidance for selective Gaussian densification.

Keywords: 3D Gaussian Splatting, novel view synthesis, Gaussian densification, density control, perceptual guidance, structural guidance

1 Introduction

Novel-view synthesis seeks to render photorealistic images from previously unobserved viewpoints using a set of posed input images. Neural radiance fields (NeRFs) represent scenes as continuous implicit functions and have demonstrated high rendering quality [1]. However, their reliance on dense ray sampling and repeated network evaluation often entails substantial computational cost during training and rendering. Three-dimensional Gaussian Splatting (3DGS) provides an explicit alternative by representing a scene with anisotropic Gaussian primitives and rendering them through differentiable splatting and alpha compositing [2]. This representation supports efficient optimization and real-time rendering, but its reconstruction quality depends strongly on how Gaussian primitives are allocated and refined during training.

Densification is a central component of this refinement process. In the original 3DGS framework, primitives with sufficiently large accumulated view-space positional gradients are selected for densification, after which their spatial scales determine whether they are cloned or split [2]. Gradient averaging may nevertheless attenuate strong optimization signals arising from different pixels or views, resulting in insufficient refinement of fine textures, object boundaries, and thin structures. Subsequent methods have addressed this limitation through absolute-gradient accumulation, pixel-aware gradient estimation, budget-constrained allocation, resource-aware optimization, stochastic density control, and perceptual evidence modeling [3–10]. These developments improve the selection or allocation of Gaussian primitives, but the individual stages of densification are often guided by different criteria or treated separately.

Effective densification involves more than identifying primitives that require additional capacity. Once a Gaussian has been selected, the method must determine whether cloning or splitting is more appropriate and, when splitting is performed, how the child primitives should be positioned. These decisions depend on scene attributes that are not fully described by a single optimization signal. Perceptual responses identify image regions whose appearance variations strongly affect visual similarity, whereas multi-view structural evidence characterizes boundaries, thin structures, occlusion transitions, and local geometric changes. The two types of evidence are complementary but not interchangeable: a texture-rich region may exhibit a strong perceptual response without a pronounced structural discontinuity, while a low-texture contour may remain structurally important despite a weaker perceptual response. A unified primitive-level representation of these cues may therefore provide more informative guidance for successive densification decisions.

Based on this observation, we propose PSCD-GS, a role-aware perception-structure collaborative densification framework for 3DGS. Image-space perceptual responses and multi-view structural cues are first mapped to individual Gaussian primitives through rendering-contribution-weighted aggregation. The resulting Gaussian-level perceptual and structural evidence scores are converted into continuous memberships in four roles: perception-dominant, structure-dominant, co-dominant, and low-response. In contrast to a hard assignment, these memberships retain the gradual transitions between regional characteristics and allow a primitive to respond to

both types of evidence. The derived role representation is then used to coordinate candidate prioritization and to construct competing cloning and splitting utilities, while the associated response directions guide the placement of newly generated Gaussians.

Perceptually important primitives receive additional representation capacity, whereas structurally important primitives are refined with reference to local structural evidence. Low-response primitives are treated conservatively to limit unnecessary growth of the representation.

We evaluate PSCD-GS on Mip-NeRF 360 [11], Deep Blending [12], and Tanks & Temples [13]. All baseline methods are re-evaluated using their official implementations and recommended configurations on the same scene sets and train/test splits, with all experiments conducted on the same hardware platform. Method-specific image-resolution policies and training settings follow the corresponding official implementations, while evaluation metrics are computed using a common protocol. The reported results show that PSCD-GS achieves competitive or leading dataset-level image-quality averages while maintaining a controlled number of Gaussian primitives. Qualitative comparisons further indicate clearer texture details, more coherent object boundaries, and better preservation of thin structures.

The main contributions of this work are summarized as follows:

- We formulate Gaussian densification in a joint perception–structure space and derive continuous primitive-level memberships for perception-dominant, structure-dominant, co-dominant, and low-response roles. This formulation preserves the complementary and potentially overlapping characteristics of perceptual and structural evidence.
- We construct a rendering-contribution-weighted aggregation mechanism that maps image-space perceptual responses and multi-view structural cues to Gaussian primitives. The resulting role memberships support collaborative candidate prioritization and define competing cloning and splitting utilities, while the associated response directions guide the structure-aware placement of child Gaussians.
- We evaluate the proposed framework on three real-world novel-view-synthesis benchmarks. The quantitative and qualitative results show that PSCD-GS provides a favorable balance between rendering quality and representation size, with visible improvements in texture-rich and structurally challenging regions.

2 Related work

2.1 3D Gaussian Splatting and density control

3D Gaussian Splatting (3DGS) represents a scene using anisotropic Gaussian primitives and renders novel views through differentiable splatting and alpha compositing [2]. Its explicit representation enables efficient optimization and real-time rendering, but reconstruction quality depends on the spatial distribution, scale, and number of Gaussian primitives. Several subsequent studies have improved other components of the 3DGS pipeline. Mip-Splatting introduces three-dimensional smoothing and two-dimensional filtering to reduce aliasing [14]; GaussianPro progressively propagates Gaussian primitives into poorly initialized regions using geometric priors [15];

and Scaffold-GS organizes local neural Gaussians around anchor points to support view-adaptive rendering [16]. These approaches improve representation quality or organization, but they do not primarily investigate how multiple regional cues should coordinate the densification decisions considered in this work.

The original 3DGS adaptive density control selects primitives according to accumulated positional gradients in the view-space and applies cloning or splitting according to their spatial scales [2]. Subsequent methods have revised different parts of this procedure. AbsGS accumulates absolute gradients to prevent opposing pixel-level signals from canceling each other [3], while Pixel-GS introduces a pixel-aware gradient that accounts for the number of pixels covered by each primitive [4]. Mini-Splatting constrains the number of Gaussians through densification, simplification, and importance-based sampling [5], while Taming 3DGS regulates primitive growth under limited computational resources [6]. The Monte Carlo Markov chain formulation of 3DGS replaces heuristic density control with stochastic relocation and growth operations [7]. SteepGS instead interprets density control from an optimization perspective and seeks a compact distribution through steepest-descent updates [8].

More recent work has extended densification beyond candidate scoring. Deng et al. [10] combine an edge-aware candidate score with long-axis splitting, recovery-aware pruning, growth control, and multi-view updates, thereby addressing candidate selection, primitive transformation, and overfitting within a broader density-control pipeline. These developments show that effective densification requires coordinated treatment of several stages rather than an isolated selection criterion. PSCD-GS shares this general motivation, but focuses on deriving a common primitive-level representation from complementary perceptual and structural evidence.

2.2 Perceptual- and structure-aware densification

Image-domain priors have increasingly been used to identify regions that require additional Gaussian capacity. The learned perceptual features provide a measure of visual similarity that is more closely related to human judgments than conventional pixel-wise distances [17]. Perceptual-GS incorporates scene-adaptive perceptual sensitivity into Gaussian allocation and assigns finer granularity to visually important regions while constraining the representation size [9]. CDC-GS introduces a loss-agnostic visual-complexity prior derived from high-frequency wavelet components and encourages consistency between image complexity and local primitive density; this consistency is then used to identify primitives for densification and pruning [18]. These methods demonstrate that appearance complexity and perceptual sensitivity can provide useful information beyond rendering gradients or pixel-wise residuals.

Structural evidence offers a related but distinct source of guidance. Surface-oriented representations, including SuGaR [19], 2D Gaussian Splatting [20], and Gaussian Opacity Fields [21], improve geometric alignment or surface reconstruction by imposing stronger structural constraints on the Gaussian representation. Although their objectives differ from density control, they demonstrate the importance of primitive orientation and spatial organization for representing scene geometry. More directly, SAD-GS estimates multi-scale image structure using structure tensors and Laplacian

scale-space analysis. It compares projected Gaussian extents with local image frequencies and performs multi-view-consistent, axis-dependent anisotropic splitting when a primitive under-resolves the observed structure [22]. Deng et al. [10] likewise use edge information to identify densification candidates and employ long-axis splitting to reduce geometric distortion.

Perceptual and structural evidence nevertheless describe different properties of a scene. Perceptual or frequency-based scores emphasize appearance-sensitive and texture-rich regions, whereas structural cues can highlight boundaries, occlusion transitions, thin objects, and local geometric changes. A region may exhibit a strong response in one dimension but a weak response in the other, and the two cues may also be simultaneously important. PSCD-GS therefore maps both types of evidence to individual Gaussian primitives and represents their relationship through continuous perception-dominant, structure-dominant, co-dominant, and low-response memberships. Unlike methods based on a single complexity, perceptual, edge, or frequency score, the proposed framework retains the complementary and overlapping responses of the two cue types. The resulting memberships are used to compute a collaborative densification-priority score and to construct competing role-conditioned cloning and splitting utilities, while the corresponding evidence directions guide the placement of child Gaussians.

3 Methods

3.1 Problem formulation and overview

Three-dimensional Gaussian Splatting (3DGS) represents a scene by a set of anisotropic Gaussian primitives $\mathcal{G} = \{g_i\}_{i=1}^N$. Each primitive is parameterized by a three-dimensional center $\boldsymbol{\mu}_i$, an anisotropic covariance $\boldsymbol{\Sigma}_i$, an opacity, and spherical-harmonic appearance coefficients. These parameters are jointly optimized through differentiable splatting and alpha compositing [2]. In addition to updating the primitive parameters, 3DGS periodically adapts the number and spatial distribution of Gaussians through densification and pruning.

The original adaptive density control uses the accumulated view-space positional gradient to identify primitives that require additional representation capacity. Let $\mathcal{T}_i$ be the set of statistic-update steps at which Gaussian g_i is visible. Its average gradient magnitude and maximum spatial scale are written as

$$\bar{g}_i = \frac{1}{|\mathcal{T}_i| + \epsilon} \sum_{t \in \mathcal{T}_i} \left\| \frac{\partial \mathcal{L}_t}{\partial \boldsymbol{\mu}_{i,t}^{2D}} \right\|_2, \quad s_i^{\max} = \|\mathbf{s}_i\|_\infty, \quad (1)$$

where $\mathcal{L}_t$ denotes the rendering loss and $\mathbf{s}_i$ is the activated scale vector. A primitive becomes eligible for densification when $\bar{g}_i$ exceeds the gradient threshold τ_g . The baseline then clones a spatially small candidate and splits a spatially large candidate according to a scene-scale-dependent threshold τ_s . Cloning duplicates the selected primitive at the same initial center, whereas splitting replaces the parent with smaller

children sampled from its covariance support. The resulting primitives continue to be updated by the original optimization process.

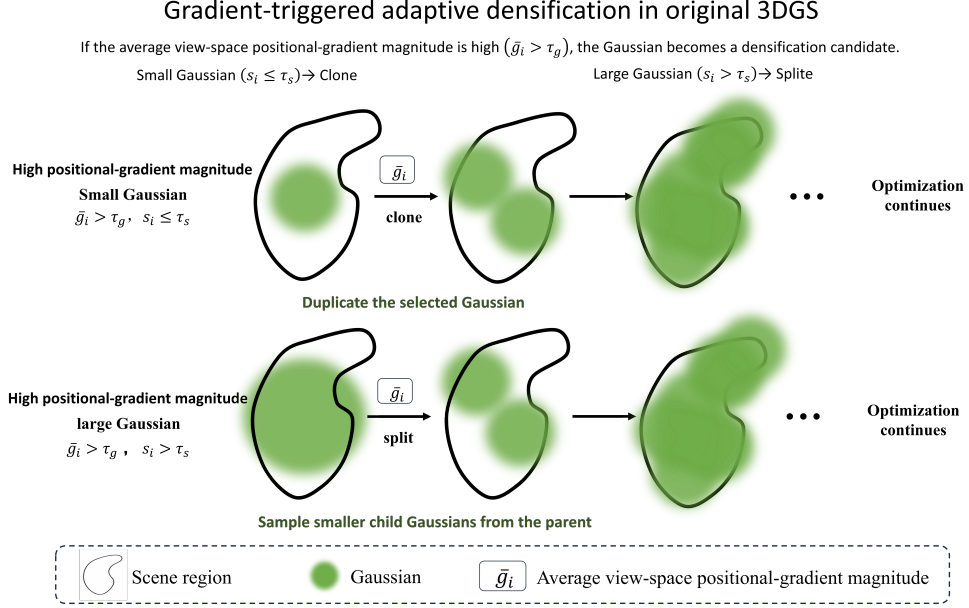


Fig. 1 Adaptive density control in the original 3DGS. The accumulated magnitude of the view-space positional gradient determines which Gaussians are eligible for densification, while the maximum spatial scale determines whether cloning or splitting is applied. Cloning duplicates a small Gaussian at the same initial center. Splitting replaces a large parent with smaller children sampled from its covariance support, after which all primitives continue to be optimized

This rule provides two useful signals: the gradient measures the current optimization pressure, and the scale prevents an oversized primitive from being duplicated without spatial decomposition. However, neither signal describes the image-space characteristics associated with the primitive. A large positional gradient does not indicate whether the unresolved variation arises from appearance detail, a structural boundary, or both. Likewise, the scale-based operation rule does not determine whether a small primitive requires additional appearance capacity or an explicit spatial separation. The baseline split initialization also does not use the direction of the image evidence. Consequently, Gaussians with similar gradients and scales receive the same densification treatment even when they represent different regional conditions.

PSCD-GS extends adaptive density control by introducing a continuous perception-structure role representation. At a densification event k , the method takes the training views and the current Gaussian set $\mathcal{G}_k$ as input. It first constructs perceptual and structural evidence in the training views and transfers this evidence to the visible Gaussian primitives. The resulting Gaussian-level scores describe the strength

of the two evidence types and are converted into continuous co-dominant, perception-dominant, structure-dominant, and low-response roles. These roles then condition three successive decisions: which Gaussians should be densified, whether each selected Gaussian should be cloned or split, and how reliable directional evidence should be used to initialize split children.

The proposed framework does not replace the Gaussian representation, differentiable renderer, reconstruction objective, or the general optimization and pruning schedule of 3DGS. The original positional gradient remains an explicit measure of optimization demand, and the original scale criterion is retained as a safeguard for oversized primitives. PSCD-GS instead augments these baseline signals with Gaussian-level perception–structure evidence and uses different role-conditioned read-outs for candidate selection, operation selection, and split placement. This separation is important because the three decisions answer different questions: whether additional capacity is needed, whether that capacity should preserve or redistribute spatial support, and where the redistributed support should be initialized.

3.2 Perceptual and structural evidence

The first stage constructs two complementary response fields from the training views. Its input is a posed training image I^v and, for structural cues, the current depth rendered from $\mathcal{G}_k$ at densification event k . Its output is a perceptual magnitude map A^v , a structural magnitude map B_k^v , and their corresponding image-plane direction tensors. Perceptual evidence identifies image regions with pronounced multi-scale feature variation, whereas structural evidence describes boundaries and geometric discontinuities that are supported by the current multi-view reconstruction. The two branches are kept separate because a textured but geometrically smooth surface and a weakly textured structural boundary can require different forms of refinement. Throughout this section, $\text{RNorm}_{\Omega^v}(\cdot)$ denotes robust image-space normalization: the response is clipped at fixed lower and upper quantiles within view v and then linearly mapped to $[0, 1]$.

3.2.1 Perceptual response

For each training image I^v , a fixed ImageNet-pretrained VGG-16 network extracts the channel-normalized feature maps used by the LPIPS representation [17]. The network remains frozen. We measure the spatial variation of multiple feature levels and resize the responses to the evidence resolution:

$$\begin{aligned} A^v &= \text{RNorm}_{\Omega^v} \left[\sum_{l \in \mathcal{L}} \beta_l \text{Up}_l \left(\left\| \nabla \tilde{\phi}_l(I^v) \right\|_2 \right) \right], \\ \mathbf{R}_A^v(\mathbf{u}) &= \sum_{l \in \mathcal{L}} \beta_l \text{Up}_l \left[\sum_c \nabla \tilde{\phi}_{l,c}^v(\mathbf{u}) (\nabla \tilde{\phi}_{l,c}^v(\mathbf{u}))^\top \right]. \end{aligned} \quad (2)$$

Here, $\mathcal{L}$ is the set of selected layers, β_l is the corresponding layer weight, and Up_l denotes spatial resizing. The scalar map A^v measures the strength of local perceptual

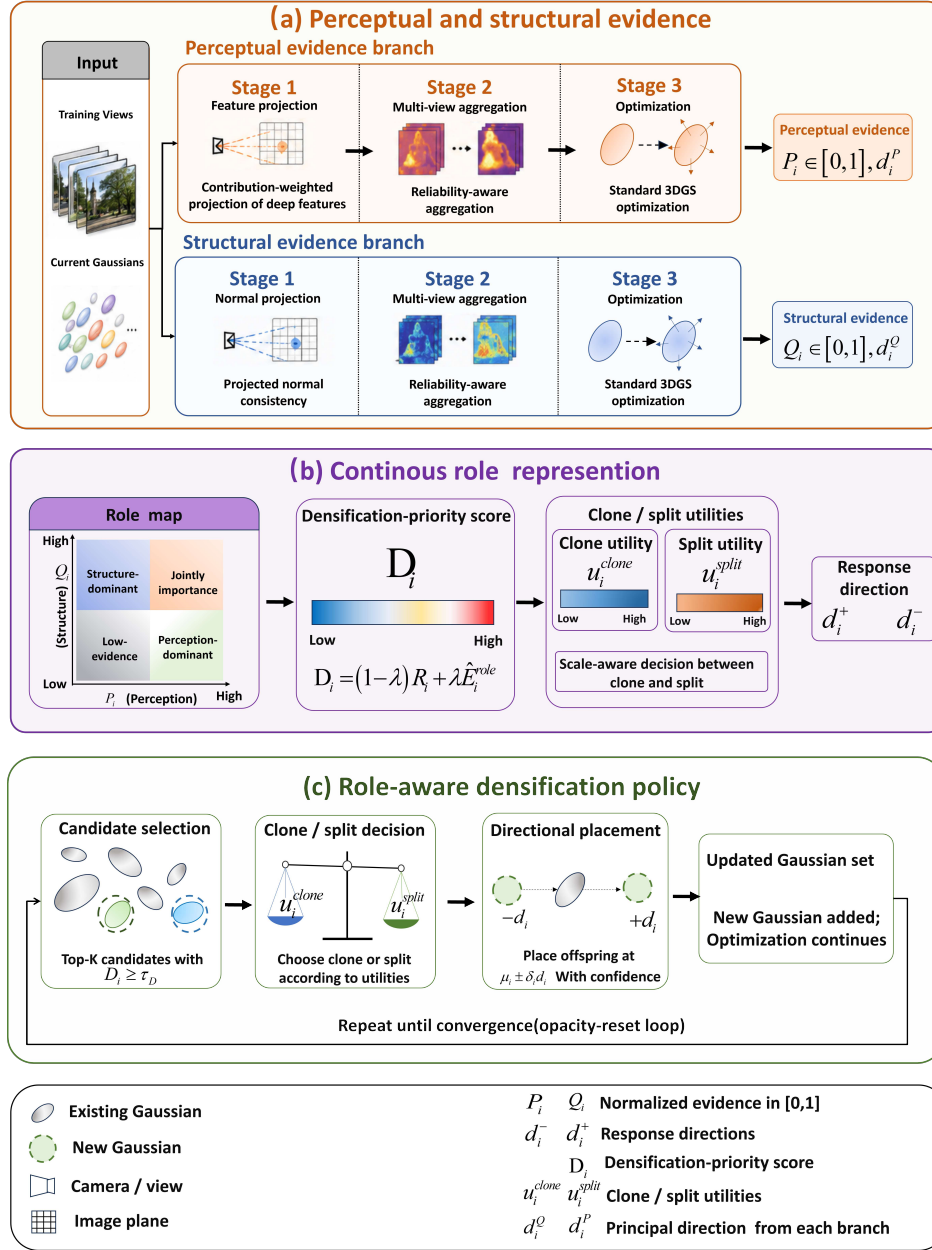


Fig. 2 Overview of PSCD-GS. Perceptual and structural evidence are constructed from the training views and the current Gaussian set, aggregated to Gaussian primitives through rendering contributions, converted into continuous perception–structure roles, and used to coordinate candidate prioritization, clone-or-split selection, and confidence-aware split placement. The generated primitives are then returned to the original 3DGS optimization, pruning, and opacity-reset loop

variation. The symmetric tensor $\mathbf{R}_A^v$ retains its dominant image-plane axis without assigning a sign to that axis. Because both quantities depend only on the training images and the frozen feature extractor, they are computed once before optimization.

3.2.2 Structural response and cross-view reliability

Structural evidence combines a fixed image-boundary cue with a geometry-dependent depth-discontinuity cue. Let Y^v be the luminance image and let Z_k^v be the depth rendered from the current Gaussian set at densification event k . The image response E^v is obtained from the luminance gradient, while D_k^v uses the gradient of logarithmic depth so that relative, rather than absolute, depth changes are measured. Their weighted sum forms a single-view structural response H_k^v .

A strong single-view response is not necessarily a reliable structural observation. Texture edges, erroneous depth, and occlusion can all produce responses that do not persist under reprojection. We therefore interpret multi-view consistency as reliability rather than as an additional structural magnitude. For each view v , a fixed set $\mathcal{N}(v)$ of overlapping neighboring views is selected from the camera geometry. A valid pixel $\mathbf{u}$ is back-projected with $Z_k^v(\mathbf{u})$ and reprojected to a neighbor v' , yielding $\mathbf{u}'$ and the transferred depth $z_{v \rightarrow v'}$. The complete structural response is

$$\begin{aligned}
E^v &= \text{RNorm}_{\Omega^v}(\|\nabla Y^v\|_2), \\
D_k^v &= \text{RNorm}_{\Omega^v}(\|\nabla \log(Z_k^v + \epsilon_z)\|_2), \\
H_k^v &= \lambda_e E^v + \lambda_d D_k^v, \quad \lambda_e + \lambda_d = 1, \\
\Delta_z^{vv'}(\mathbf{u}) &= \frac{|Z_k^{v'}(\mathbf{u}') - z_{v \rightarrow v'}|}{\max[Z_k^{v'}(\mathbf{u}'), z_{v \rightarrow v'}, \epsilon_z]}, \\
a_{vv'}(\mathbf{u}) &= \chi_{vv'}(\mathbf{u}) \exp\left(-\frac{\Delta_z^{vv'}(\mathbf{u})}{\sigma_z}\right) \exp\left(-\frac{|H_k^v(\mathbf{u}) - \tilde{H}_k^{v'}(\mathbf{u}')|}{\sigma_h}\right), \\
M_k^v(\mathbf{u}) &= \frac{\sum_{v' \in \mathcal{N}(v)} a_{vv'}(\mathbf{u})}{\sum_{v' \in \mathcal{N}(v)} \chi_{vv'}(\mathbf{u}) + \epsilon}, \\
B_k^v &= \text{RNorm}_{\Omega^v}(H_k^v [m_{\min} + (1 - m_{\min})M_k^v]).
\end{aligned} \tag{3}$$

The binary variable $\chi_{vv'}(\mathbf{u})$ is one only when the reprojection is inside the neighboring image, both depths are valid, and $\Delta_z^{vv'}(\mathbf{u}) < \tau_z$. Otherwise it equals zero. The locally pooled response $\tilde{H}_k^{v'}$ tolerates small reprojection errors. If no valid correspondence exists, $M_k^v(\mathbf{u})$ is set to zero. The floor $m_{\min} > 0$ prevents genuine occlusion boundaries and thin structures from being completely removed when they have limited cross-view support. Thus, M_k^v only determines how strongly the single-view response should be trusted; it does not create structural evidence in a region where H_k^v is weak.

For later directional analysis, let $\bar{\mathbf{g}}_E^v$ and $\bar{\mathbf{g}}_{D,k}^v$ denote the luminance- and log-depth-gradient vectors divided by the same robust scales used to construct E^v and

D_k^v . The structural branch stores the symmetric tensor

$$\mathbf{R}_{B,k}^v(\mathbf{u}) = [m_{\min} + (1 - m_{\min})M_k^v(\mathbf{u})] [\lambda_e \bar{\mathbf{g}}_E^v(\mathbf{u}) \bar{\mathbf{g}}_E^v(\mathbf{u})^\top + \lambda_d \bar{\mathbf{g}}_{D,k}^v(\mathbf{u}) \bar{\mathbf{g}}_{D,k}^v(\mathbf{u})^\top]. \quad (4)$$

The tensor describes the dominant image-plane axis of the accepted structural response, while B_k^v describes its magnitude. The fixed boundary component is pre-computed, whereas depth, cross-view reliability, and the final structural response are refreshed at each densification event. The formulation uses only the two structural mixture weights in Eq. (3), namely λ_e and λ_d , with $\lambda_e + \lambda_d = 1$.

3.3 Continuous perception–structure role representation

The second stage maps the view-space evidence to the primitives that generated the rendered pixels. Its input is the pair of evidence maps (A^v, B_k^v) and the alpha-compositing contribution of each visible Gaussian. Its output is a pair of calibrated Gaussian-level scores, P_i and Q_i , followed by a four-component continuous role vector. Sampling a response only at the projected Gaussian center is insufficient when a projected footprint spans several pixels or overlaps other primitives. We therefore use the same alpha-compositing contributions as the renderer and aggregate the two evidence branches independently.

3.3.1 Contribution-weighted Gaussian aggregation

For pixel $\mathbf{u}$ in view v , the rendering contribution of Gaussian g_i is $w_i^v(\mathbf{u}) = T_i^v(\mathbf{u})\alpha_i^v(\mathbf{u})$, where T_i^v is the transmittance accumulated before g_i and α_i^v is its effective alpha value. For either view-space map $X_k^v \in \{A^v, B_k^v\}$, we compute

$$C_i^v = \sum_{\mathbf{u} \in \Omega^v} w_i^v(\mathbf{u}), \quad x_i^{X,v} = \frac{\sum_{\mathbf{u} \in \Omega^v} w_i^v(\mathbf{u}) X_k^v(\mathbf{u})}{C_i^v + \epsilon}. \quad (5)$$

The total contribution C_i^v measures how strongly the primitive is observed in view v , and $x_i^{X,v}$ is its contribution-weighted response in that view. These quantities can be accumulated during compositing without storing the dense contribution tensor.

Evidence observed in only one weak view should not be treated in the same way as evidence supported by several views. We convert C_i^v into a bounded visibility weight ν_i^v , use the normalized weights to estimate the cross-view response and its dispersion, and attenuate observations with insufficient support or large disagreement:

$$\begin{aligned} \nu_i^v &= \min\left(1, \frac{C_i^v}{\tau_C}\right), & \omega_i^v &= \frac{\nu_i^v}{\sum_{v'} \nu_i^{v'} + \epsilon}, \\ \mu_i^X &= \sum_v \omega_i^v x_i^{X,v}, & V_i^X &= \sum_v \omega_i^v (x_i^{X,v} - \mu_i^X)^2, \\ \tilde{I}_i^X &= \min\left(1, \frac{\sum_v \nu_i^v}{K_{\text{vis}}}\right) \exp\left(-\frac{V_i^X}{\sigma_X^2 + \epsilon}\right) \mu_i^X. \end{aligned} \quad (6)$$

The first factor in $\tilde{I}_i^X$ measures the amount of view support, the exponential term measures cross-view agreement, and μ_i^X preserves the response strength. Unlike a coefficient of variation, the bounded variance penalty remains stable when the mean response is close to zero. If a Gaussian has no reliable view, its raw evidence is set to zero.

The distributions of raw Gaussian evidence change as primitives are added, removed, and reshaped. To prevent event-wise extrema from changing the score scale abruptly, lower and upper calibration quantiles are updated by an exponential moving average:

$$\begin{aligned} q_{p,k}^X &= \beta_q q_{p,k-1}^X + (1 - \beta_q) \text{Quantile}_p \left(\{\tilde{I}_i^X\}_{i \in \mathcal{G}_k} \right), \\ \text{Cal}_k^X(x) &= \text{clip} \left(\frac{x - q_{\text{low},k}^X}{q_{\text{high},k}^X - q_{\text{low},k}^X + \epsilon_{\text{cal}}}, 0, 1 \right), \\ P_i &= \text{Cal}_k^A(\tilde{I}_i^A), & Q_i &= \text{Cal}_k^B(\tilde{I}_i^B). \end{aligned} \tag{7}$$

Here, P_i and Q_i are the Gaussian-level perceptual and structural evidence scores. If the current quantile span collapses, the previous valid calibration range is retained; before a valid range is available, the corresponding branch is set to zero. This rule prevents a nearly constant response field from being amplified into artificial evidence.

3.3.2 Continuous perception–structure roles

Directly summing P_i and Q_i would indicate overall response strength but would discard how that response is distributed between the two evidence types. We instead map the calibrated scores to smooth memberships and factorize their joint state into four continuous roles:

$$\begin{aligned} m_i^P &= \sigma[a_P(P_i - \tau_P)], & m_i^Q &= \sigma[a_Q(Q_i - \tau_Q)], \\ \rho_i^{PQ} &= m_i^P m_i^Q, & \rho_i^P &= m_i^P (1 - m_i^Q), \\ \rho_i^Q &= (1 - m_i^P) m_i^Q, & \rho_i^0 &= (1 - m_i^P)(1 - m_i^Q). \end{aligned} \tag{8}$$

The four weights represent co-dominant, perception-dominant, structure-dominant, and low-response behavior and satisfy $\rho_i^{PQ} + \rho_i^P + \rho_i^Q + \rho_i^0 = 1$. They are not hard semantic labels. A Gaussian can move continuously between roles as its evidence changes during optimization. The roles provide the shared state for subsequent densification decisions, but they do not remove the need for gradient demand, scale safeguards, or directional confidence.

3.4 Role-conditioned densification

The role representation is consumed through three distinct readouts. Candidate selection treats perceptual and structural evidence symmetrically because either can indicate a need for additional capacity. Operation selection is deliberately asymmetric

because structural evidence more directly supports spatial decomposition. Split placement additionally depends on whether the image-plane response axes can be lifted and fused reliably. Separating the readouts prevents one score from being forced to answer three different questions.

3.4.1 Candidate priority and growth control

The perceptual and structural roles describe the potential value of refinement, whereas the original positional gradient measures whether the current reconstruction is still exerting optimization pressure on the primitive. We retain both quantities. The raw gradient is mapped relative to the original threshold, and the role-weighted evidence score treats the two single-dominant roles symmetrically:

$$\begin{aligned}
G_i &= \frac{\bar{g}_i}{\bar{g}_i + \tau_g + \epsilon}, \\
S_i &= \rho_i^P P_i + \rho_i^Q Q_i + \rho_i^{PQ} \frac{P_i + Q_i}{2}, \\
D_i &= G_i + \alpha(1 - G_i)S_i, \\
K_k &= \max(0, \min[K_{\max}, \lfloor r_{\text{grow}} N_k \rfloor, N_{\max} - N_k]), \\
C_k &= \text{TopK}_{K_k} \{i : D_i \geq \tau_D, i \notin \mathcal{P}_k\}.
\end{aligned} \tag{9}$$

Here, TopK_{K_k} selects up to K_k eligible primitives in descending order of D_i . The quantity $G_i \in [0, 1]$ preserves the ordering of the original gradient statistic, $\alpha \in [0, 1]$ controls the maximum evidence bonus, and $\mathcal{P}_k$ contains primitives already scheduled for pruning. The gradient-anchored form ensures that evidence increases, rather than replaces, current optimization demand. The low-response role contributes no direct evidence term, but a low-response Gaussian can still be selected when its gradient is sufficiently large.

The event budget K_k limits both the fraction of the current representation that can be densified and the absolute number of candidates. It also enforces a global cap $N_{\max}$. With one clone per cloned parent and two children replacing each split parent, both operations increase the representation by one primitive before pruning. Candidate selection therefore uses the same net-growth budget for clone and binary split. If fewer than K_k primitives pass τ_D , the remaining budget is not filled.

3.4.2 Clone-or-split decision

Candidate priority answers where additional capacity is useful, but it does not specify how that capacity should be introduced. For a small perception-dominant Gaussian, cloning is the conservative operation because it adds local appearance and opacity degrees of freedom without immediately separating the spatial support. A structure-dominant Gaussian is more compatible with splitting because its evidence indicates a boundary or geometric change across which a single support may be inadequate. A co-dominant Gaussian receives a base preference for splitting, modulated by the structural share of its evidence. Low-response candidates fall back to the original size-based behavior.

We encode this policy with a bounded splitting tendency and retain the original scale safeguard:

$$\begin{aligned}\zeta_i^Q &= \frac{Q_i}{P_i + Q_i + \epsilon}, \\ R_i^{\text{split}} &= \rho_i^Q + \rho_i^{PQ} \left[\kappa_{PQ} + (1 - \kappa_{PQ}) \zeta_i^Q \right], \\ o_i &= \begin{cases} \text{split}, & s_i^{\max} > \tau_s \vee R_i^{\text{split}} \geq \tau_r, \\ \text{clone}, & \text{otherwise.} \end{cases}\end{aligned}\tag{10}$$

The coefficient $\kappa_{PQ} \in [0, 1]$ specifies the base split preference of the co-dominant role, and τ_r converts the continuous tendency into an operation. A spatially large Gaussian is always split, irrespective of its role, because cloning would duplicate the same oversized support. The role mapping only changes the operation assigned to candidates below the scale threshold. This mapping is an explicit optimization hypothesis and must therefore be evaluated separately from candidate selection.

3.4.3 Confidence-aware split placement

Directional evidence is evaluated only for candidates assigned to split. For compact notation, let $\mathcal{X}_P^v = A^v$ and $\mathcal{X}_Q^v = B_k^v$, with corresponding image-plane tensors $\mathbf{R}_P^v = \mathbf{R}_A^v$ and $\mathbf{R}_Q^v = \mathbf{R}_{B,k}^v$. For branch $X \in \{P, Q\}$, the contribution-weighted tensor of Gaussian g_i in view v is

$$\mathbf{T}_i^{X,v} = \frac{\sum_{\mathbf{u} \in \Omega^v} w_i^v(\mathbf{u}) \mathcal{X}_X^v(\mathbf{u}) \mathbf{R}_X^v(\mathbf{u})}{\sum_{\mathbf{u} \in \Omega^v} w_i^v(\mathbf{u}) \mathcal{X}_X^v(\mathbf{u}) + \epsilon}.\tag{11}$$

Let $\lambda_{1,i}^{X,v} \geq \lambda_{2,i}^{X,v} \geq 0$ be the eigenvalues of $\mathbf{T}_i^{X,v}$ and $\mathbf{e}_i^{X,v}$ its leading unit eigenvector. The tensor uses outer products, so $\mathbf{e}$ and $-\mathbf{e}$ describe the same response axis.

A two-dimensional response axis does not uniquely determine a three-dimensional surface direction. We therefore seek a displacement inside the support of the current Gaussian whose projection best reproduces the image-plane axis. Let $\boldsymbol{\Sigma}_i = \mathbf{L}_i \mathbf{L}_i^\top$ and let $\mathbf{J}_i^v$ be the 2×3 projection Jacobian at $\boldsymbol{\mu}_i$. The covariance-regularized lift is

$$\begin{aligned}\overline{\mathbf{A}}_i^v &= \frac{\mathbf{J}_i^v \mathbf{L}_i}{\|\mathbf{J}_i^v \mathbf{L}_i\|_F + \epsilon}, \\ \mathbf{z}_i^{X,v} &= \left(\overline{\mathbf{A}}_i^{v\top} \overline{\mathbf{A}}_i^v + \lambda_J \mathbf{I}_3 \right)^{-1} \overline{\mathbf{A}}_i^{v\top} \mathbf{e}_i^{X,v}, \\ \mathbf{t}_i^{X,v} &= \text{normalize} \left(\mathbf{L}_i \mathbf{z}_i^{X,v} \right).\end{aligned}\tag{12}$$

This is a ridge-regularized solution in whitened Gaussian coordinates. The resulting $\mathbf{t}_i^{X,v}$ is a Gaussian-support-constrained response-displacement axis. It is not interpreted as a true surface tangent or normal.

The lift is used only when its source tensor is anisotropic, the Gaussian has sufficient visibility and response support, and the projected displacement fits the observed image-plane axis. We define the single-view confidence and its sign-invariant

multi-view fusion as

$$\begin{aligned}
a_i^{X,v} &= \frac{\lambda_{1,i}^{X,v} - \lambda_{2,i}^{X,v}}{\lambda_{1,i}^{X,v} + \lambda_{2,i}^{X,v} + \epsilon}, \\
\xi_i^{X,v} &= \min\left(1, \frac{\sum_{\mathbf{u}} w_i^v(\mathbf{u}) \mathcal{X}_X^v(\mathbf{u})}{\tau_X}\right), \\
f_i^{X,v} &= \exp\left[-\frac{\|\bar{\mathbf{A}}_i^v \mathbf{z}_i^{X,v} - \mathbf{e}_i^{X,v}\|_2^2}{\sigma_{\text{fit}}^2}\right], \\
r_i^{P,v} &= 1, \quad r_i^{Q,v} = \frac{\sum_{\mathbf{u}} w_i^v(\mathbf{u}) M_k^v(\mathbf{u})}{C_i^v + \epsilon}, \\
c_i^{X,v} &= a_i^{X,v} \nu_i^v \xi_i^{X,v} f_i^{X,v} r_i^{X,v}, \\
\mathbf{F}_i^X &= \frac{\sum_v c_i^{X,v} \mathbf{t}_i^{X,v} \mathbf{t}_i^{X,v\top}}{\sum_v c_i^{X,v} + \epsilon}.
\end{aligned} \tag{13}$$

All confidence factors lie in $[0, 1]$. The structural reliability $r_i^{Q,v}$ is the contribution-weighted mean of the cross-view gate, while the perceptual branch uses $r_i^{P,v} = 1$. Let $\Lambda_{1,i}^X \geq \Lambda_{2,i}^X \geq \Lambda_{3,i}^X$ be the eigenvalues of $\mathbf{F}_i^X$. The branch direction and confidence are

$$\mathbf{d}_i^X = \text{eig}_{\max}(\mathbf{F}_i^X), \quad c_i^X = \frac{\Lambda_{1,i}^X - \Lambda_{2,i}^X}{\Lambda_{1,i}^X + \epsilon} \min\left(1, \frac{\sum_v c_i^{X,v}}{K_{\text{dir}}}\right). \tag{14}$$

The eigenvalue gap measures whether the lifted axes agree on a dominant three-dimensional axis, and the second factor requires sufficient view support. A branch with $c_i^X < \tau_{\text{dir}}$ is excluded from placement.

The reliable branch axes are finally combined using the soft memberships rather than the candidate or operation scores:

$$\begin{aligned}
\tilde{c}_i^X &= c_i^X \mathbb{I}[c_i^X \geq \tau_{\text{dir}}], \\
\mathbf{H}_i^{\text{role}} &= m_i^P \tilde{c}_i^P \mathbf{d}_i^P \mathbf{d}_i^{P\top} + m_i^Q \tilde{c}_i^Q \mathbf{d}_i^Q \mathbf{d}_i^{Q\top}, \\
\mathbf{d}_i &= \text{eig}_{\max}(\mathbf{H}_i^{\text{role}}), \\
\delta_i &= \eta_{\delta} \sqrt{\mathbf{d}_i^{\top} \mathbf{\Sigma}_i \mathbf{d}_i}, \quad \boldsymbol{\mu}_i^{\pm} = \boldsymbol{\mu}_i \pm \delta_i \mathbf{d}_i.
\end{aligned} \tag{15}$$

The combined tensor weights the perceptual and structural axes according to their soft memberships and directional confidences. In a perception-dominant or structure-dominant region, the leading axis is mainly determined by the corresponding branch. In a co-dominant region, both branches contribute to the fused tensor, from which a single dominant placement axis is selected only when the fused evidence is sufficiently anisotropic. When the two branches provide conflicting or nearly isotropic directional evidence, the resulting low eigenvalue gap triggers the fallback strategy described

below. Let $\Gamma_{1,i} \geq \Gamma_{2,i} \geq \Gamma_{3,i}$ be its eigenvalues. Its confidence is defined as $c_i^{\text{role}} = (\Gamma_{1,i} - \Gamma_{2,i}) / (\Gamma_{1,i} + \epsilon) \min[1, m_i^P \tilde{c}_i^P + m_i^Q \tilde{c}_i^Q]$. If $c_i^{\text{role}} < \tau_{\text{dir}}$, the principal covariance axis is used only when $c_i^\Sigma = (s_{i,1}^2 - s_{i,2}^2) / (s_{i,1}^2 + \epsilon) \geq \tau_\Sigma$, where $s_{i,1} \geq s_{i,2} \geq s_{i,3}$ are the ordered activated parent scales. If the parent covariance is also nearly isotropic, PSCD-GS falls back to the original 3DGS split sampling from $\mathcal{N}(\mathbf{0}, \Sigma_i)$. This fallback hierarchy avoids imposing a precise direction when the available evidence does not support one.

For a reliable direction, Eq. (15) places two children symmetrically at a fraction η_δ of the parent standard deviation along that direction. The directional covariance radius is always bounded by the maximum parent scale and therefore avoids the excessive displacement that can result from using s_i^{max} along a short covariance axis. Following the original binary split, the activated child scales are set to $\mathbf{s}_i / (0.8N_s)$ with $N_s = 2$. The children inherit the parent rotation, opacity, and appearance coefficients, and the parent is removed. A clone instead copies the parent parameters at the same initial center. Both operations are followed by ordinary gradient-based optimization.

3.5 Integration with 3DGS optimization

PSCD-GS is inserted into the densification stage of the original training loop. Before optimization, the perceptual maps, perceptual tensors, image-boundary responses, and camera-neighbor graph are precomputed. During ordinary training iterations, the renderer and reconstruction objective are unchanged, and the original view-space positional gradients are accumulated over the current densification interval. No role decision is made at these iterations.

At each densification event, the current Gaussian parameters are treated as a synchronized snapshot. Depth is rendered at the evidence resolution for the training views, after which the depth-discontinuity maps, cross-view reliability gates, and structural tensors are refreshed. An auxiliary compositing pass accumulates C_i^v , the two contribution-weighted response sums, and the compact symmetric-tensor statistics required by Eq. (11). The accumulators are reduced across views to obtain P_i and Q_i , followed by the role weights in Eq. (8).

The event then follows the same causal order as the method description. Equation (9) selects at most K_k candidates. Equation (10) assigns clone or split while retaining the scale safeguard. Direction lifting and confidence evaluation are performed only for the selected split candidates. The resulting primitives are inserted into the Gaussian set, the selected split parents are removed, and the original opacity-, screen-size-, and world-size-based pruning rules are applied. Densification statistics and evidence accumulators are then reset, and optimization resumes with the updated Gaussian set. The original opacity-reset schedule is retained.

A complete densification event is summarized in Algorithm 1, while the default implementation hyperparameters are reported in Table 1.

The fixed perceptual branch introduces only a one-time feature-extraction cost. A structural refresh requires image-space operations and reprojection to a small number of neighboring views, with complexity linear in the number of evidence pixels and neighbors. Contribution statistics are accumulated while traversing the same splat-pixel interactions used by compositing, requiring linear auxiliary work in the number

Algorithm 1 One PSCD-GS densification event at event index k

Require: Current Gaussian set $\mathcal{G}_k$, training views, precomputed perceptual maps A^v and tensors $\mathbf{R}_A^v$, baseline gradient statistics $\bar{g}_i$, baseline pruning and opacity-reset schedules

Ensure: Updated Gaussian set after role-aware densification and baseline pruning

- 1: Render synchronized depth maps Z_k^v from $\mathcal{G}_k$
 - 2: Refresh structural maps D_k^v , M_k^v , B_k^v , and $\mathbf{R}_{B,k}^v$ using Eq. (3) and Eq. (4)
 - 3: Accumulate contribution-weighted perceptual, structural, visibility, and tensor statistics during an auxiliary compositing pass
 - 4: Compute Gaussian-level evidence scores P_i and Q_i using Eq. (5)–(7)
 - 5: Compute continuous role weights using Eq. (8)
 - 6: Select candidates $\mathcal{C}_k$ under the event-wise growth budget using Eq. (9)
 - 7: **for** each selected candidate $g_i \in \mathcal{C}_k$ **do**
 - 8: Assign clone or split using Eq. (10)
 - 9: **if** g_i is assigned to split **then**
 - 10: Estimate the branch direction and confidence using Eq. (11)–(14)
 - 11: Place split children using Eq. (15); if confidence is insufficient, use the covariance-axis or original 3DGS sampling fallback
 - 12: **else**
 - 13: Clone the parent parameters at the same initial center
 - 14: **end if**
 - 15: **end for**
 - 16: Insert new primitives, remove split parents, apply the original 3DGS pruning rules, reset event statistics, retain the opacity-reset schedule, and resume ordinary optimization
-

of processed interactions and $O(N)$ persistent storage. Role construction and candidate scoring are linear in the number of active Gaussians, whereas the 3×3 linear solves and eigendecompositions are restricted to the selected split candidates. These additional operations affect only the training-time densification stage. The final scene representation remains a standard set of 3D Gaussians and is rendered using the original rasterization pipeline.

4 Experiments

4.1 Experimental setup

4.1.1 Datasets

We evaluated PSCD-GS on three widely used real-world novel-view synthesis benchmarks: Mip-NeRF 360 [11], Tanks & Temples [13], and Deep Blending [12]. For Mip-NeRF 360, we used the five outdoor scenes *bicycle*, *flowers*, *garden*, *stump*, and *treehill*, together with the four indoor scenes *room*, *counter*, *kitchen*, and *bonsai*. For Tanks & Temples, we evaluated the *train* and *truck* scenes, while for Deep Blending, we used *Dr Johnson* and *Playroom*.

PSCD-GS followed the official 3DGS evaluation protocol. The Mip-NeRF 360 outdoor scenes used the `images_4` image directory, whereas the indoor scenes used `images_2`. Tanks & Temples and Deep Blending were evaluated using their default image directories. The official evaluation splits were used for all scenes, and no scene was excluded. All compared methods were evaluated on the same scene set and train/test partitions, while method-specific image-resolution policies and training configurations followed the corresponding official implementations.

4.1.2 Baselines

We compare PSCD-GS with the original 3DGS [2] and representative 3DGS variants, including AbsGS [3], Pixel-GS [4], Mini-Splatting [5], Taming 3DGS [6], 3DGS-MCMC [7], SteepGS [8], Perceptual-GS [9], and Deng et al. [10]. These methods cover different directions in recent 3DGS research, including gradient-based densification, pixel-aware response modeling, compact representation, sampling-based optimization, perceptual guidance, and Gaussian distribution refinement. This comparison evaluates whether the proposed perception-structure role modeling improves Gaussian allocation beyond existing density-control strategies.

4.1.3 Implementation details

PSCD-GS is implemented on top of the original 3DGS framework, with modifications restricted to the densification stage. Its rendering formulation, reconstruction losses, and general optimization pipeline follow the original 3DGS implementation. All baseline methods were reproduced from their official public repositories using the method-specific training configurations and recommended settings provided by the respective authors. This includes the official resolution policies, optimization schedules, Gaussian-budget settings, and other method-dependent parameters where applicable.

All methods were evaluated using the same dataset splits and metric implementations, and all experiments were conducted on the same workstation equipped with a single NVIDIA RTX 4090 GPU with 24 GB of memory, 16 virtual CPU cores, and 120 GB of RAM. This setup ensures that the reported comparisons are obtained under a consistent hardware and evaluation environment while preserving the officially recommended configuration of each method.

The perceptual response maps of PSCD-GS were precomputed from the training images using the fixed VGG feature extractor employed by LPIPS. The structural response maps were updated during densification using image-boundary responses, geometric discontinuities, and multi-view structural-consistency cues. The default hyperparameter settings of PSCD-GS are summarized in Table 1 and were kept fixed across all datasets and scenes.

4.1.4 Evaluation metrics

We evaluated rendering quality using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS). PSNR measures pixel-level reconstruction fidelity, SSIM evaluates structural

Table 1 Default hyperparameters of PSCD-GS. All values are fixed across datasets and scenes unless otherwise specified.

Group	Parameter	Default value
Normalization and visibility	$Q_{\text{low}}, Q_{\text{high}}$	5%, 95%
	τ_C	1×10^{-4}
	K_{vis}	3
Perceptual response	$\mathcal{L}$	relu1.2, relu2.2, and relu3.3
	β_l	1/3 for each selected feature layer
	Feature extractor	Fixed VGG-16 backbone of LPIPS; input RGB images are scaled to $[-1, 1]$, and the layer-wise response maps are bilinearly upsampled to the training-view resolution
Structural response	λ_e, λ_d	3/7, 4/7
	m_{min}	0.3
	τ_z	0.05
	σ_z, σ_h	0.05, 0.20
	$\mathcal{N}(v)$	The four training views having the largest number of shared COLMAP points with view v
Soft role assignment	a_P, a_Q	10, 10
	τ_P, τ_Q	0.5, 0.5
	β_q	0.9
	ϵ_{cal}	1×10^{-6}
	σ_X	0.25 for both evidence branches
Densification decision	α	0.4
	τ_D	0.6
	τ_s	$0.01 L_{\text{scene}}$
	τ_r	0.5
	$K_{\text{max}}, r_{\text{grow}}, N_{\text{max}}$	$5 \times 10^4, 0.05, 5 \times 10^6$
	κ_{PQ}	0.5
Split placement	λ_J	1×10^{-4}
	τ_X	1×10^{-3}
	σ_{fit}	0.25
	K_{dir}	3
	τ_{dir}	0.25
	τ_{Σ}	0.20
	η_{δ}, N_s	0.5, 2

VGG, Visual Geometry Group; LPIPS, learned perceptual image patch similarity; COLMAP, structure-from-motion and multi-view stereo reconstruction framework. Unless otherwise stated, the rendering loss, optimizer, pruning criteria, opacity-reset schedule, and the remaining training parameters follow the original 3DGS implementation.

similarity, and LPIPS measures perceptual similarity. Higher PSNR and SSIM values indicate better performance, whereas lower LPIPS values are preferred.

The same metric implementations and evaluation procedure were used for all compared methods. For each scene, the metrics were averaged over all official test views. The benchmark-level results were then calculated by averaging the scene-level scores across all included scenes. We additionally report the final number of Gaussian primitives, denoted by #G and expressed in millions, as an indicator of representation scale. The Gaussian count is reported for reference rather than treated as an independently ranked reconstruction-quality metric.

4.2 Quantitative comparison

Table 2 summarizes the quantitative comparison on Mip-NeRF 360, Deep Blending, and Tanks & Temples. Compared with the original 3DGS, PSCD-GS improves SSIM and PSNR and reduces LPIPS on all three benchmarks. These gains are consistent across pixel-level fidelity, structural similarity, and perceptual quality in the reported dataset-level averages.

The results on Mip-NeRF 360 are particularly strong. PSCD-GS ranks first in both SSIM and PSNR and second in LPIPS, with Mini-Splatting achieving the lowest LPIPS. This benchmark combines indoor and outdoor environments with dense vegetation, fine textures, irregular boundaries, and substantial viewpoint changes, making it difficult to preserve appearance details and local structures at the same time. PSCD-GS performs well under these conditions and provides a more balanced result than methods that mainly rely on gradients or pixel responses, such as AbsGS and Pixel-GS. It also achieves these gains with 2.8 million Gaussians, fewer than the 3.3 million used by the original 3DGS and substantially fewer than those used by several aggressive densification methods.

A similar balance is observed on Deep Blending. PSCD-GS obtains the highest SSIM and PSNR, while its LPIPS is second only to Perceptual-GS. Although Perceptual-GS has a slight advantage in perceptual distance, PSCD-GS reports higher SSIM and PSNR in this dataset-level comparison. Deep Blending contains indoor scenes with planar structures, door and wall boundaries, detailed textures, and view-dependent appearance effects. The results suggest that PSCD-GS can handle these different visual characteristics without favoring one aspect of reconstruction quality at the expense of the others.

Tanks & Temples presents a somewhat different pattern. Here, PSCD-GS achieves the best SSIM and LPIPS and the second-best PSNR, closely following Perceptual-GS. The scenes in this benchmark contain prominent object silhouettes, occlusion boundaries, geometric transitions, and thin structures. The improvements in SSIM and LPIPS are consistent with the intended effects of the proposed structural and perceptual cues, while the small PSNR gap indicates that these improvements are not accompanied by a substantial loss in pixel-level fidelity.

The gains of PSCD-GS are unlikely to be attributable solely to using a larger representation. The method does not aim to minimize the number of Gaussians at all costs, nor does it rely on unrestricted primitive growth. Its Gaussian count is lower than that of the original 3DGS on Mip-NeRF 360 and remains comparable

Table 2 Quantitative comparison on Mip-NeRF 360, Deep Blending, and Tanks & Temples. For PSNR, SSIM, and LPIPS, the best results are underlined and highlighted, and the second-best results are highlighted. All PSCD-GS results are boldfaced. The final number of Gaussian primitives (#G) is reported as an indicator of representation scale rather than as an independently ranked quality metric.

Method	Mip-NeRF 360				Deep Blending				Tanks & Temples			
	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	#G (M)	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	#G (M)	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	#G (M)
3DGS [2]	0.8214	27.5268	0.2047	3.3	0.9034	29.5671	0.2471	2.8	0.8495	24.5836	0.1717	1.2
AbsGS [3]	0.8346	27.6132	0.1864	3.4	0.9116	29.3384	0.2457	1.9	0.8532	23.9681	0.1795	1.3
Pixel-GS [4]	0.8317	27.7485	0.1836	5.6	0.8998	28.8749	0.2506	4.6	0.8538	23.8264	0.1703	4.5
Mini-Splatting [5]	0.8329	27.5841	<u>0.1743</u>	4.7	0.9073	29.7462	0.2308	4.6	0.8554	23.3275	0.1628	4.3
Taming 3DGS [6]	0.8238	28.0274	0.1827	3.3	0.9047	29.9865	0.2394	1.8	0.8626	24.0689	0.1654	2.8
3DGS-MCMC [7]	0.8374	28.1362	0.1905	3.3	<u>0.9172</u>	30.1647	0.2365	2.8	0.8657	24.2862	0.1596	1.2
SteepGS [8]	0.8263	28.1148	0.1926	2.3	0.9158	29.8736	0.2439	1.6	<u>0.8742</u>	24.1673	0.1635	1.3
Perceptual-GS [9]	0.8279	27.9763	0.2018	2.8	0.9085	29.9482	<u>0.2207</u>	2.9	0.8698	<u>24.6037</u>	0.1494	1.7
Deng et al. [10]	<u>0.8381</u>	<u>28.1476</u>	0.1852	1.8	0.9146	<u>30.2073</u>	0.2274	1.2	0.8736	24.5891	<u>0.1426</u>	1.2
PSCD-GS (Ours)	<u>0.8396</u>	<u>28.2439</u>	<u>0.1814</u>	2.8	<u>0.9184</u>	<u>30.2961</u>	<u>0.2243</u>	2.9	<u>0.8815</u>	<u>24.5968</u>	<u>0.1407</u>	1.3

Note: SSIM, structural similarity index measure; PSNR, peak signal-to-noise ratio; LPIPS, learned perceptual image patch similarity; #G, final number of Gaussian primitives in millions. For the reconstruction-quality metrics, $\uparrow$ indicates that higher values are better, whereas $\downarrow$ indicates that lower values are better. Best and second-best markings apply only to PSNR, SSIM, and LPIPS. The Gaussian count is provided to indicate representation scale.

to common baselines on Deep Blending and Tanks & Temples. Overall, PSCD-GS achieves improved reconstruction quality without requiring unrestricted growth of the Gaussian representation. Its final Gaussian counts remain within the range of the evaluated baselines, indicating that the quality gains are not obtained simply by using a substantially larger model.

Detailed per-scene quantitative results for all evaluated methods are provided in Online Resource 1.

4.3 Qualitative comparison

Following the quantitative comparison, we further examine the visual differences among the evaluated densification strategies. Figure 3 presents representative examples selected from different datasets and challenging scene regions, including *bicycle* and *garden* from Mip-NeRF 360, *Dr Johnson* from Deep Blending, and *truck* from Tanks & Temples. These scenes contain thin bicycle spokes, dense vegetation, indoor door frames, wall boundaries, vehicle contours, and window edges, providing representative cases for evaluating fine-detail recovery, boundary continuity, and structural fidelity.

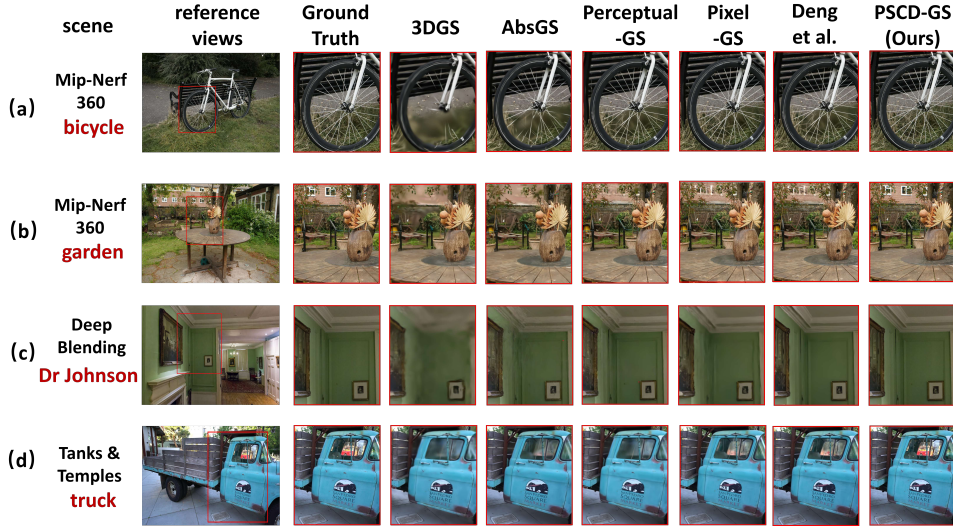


Fig. 3 Qualitative comparisons between PSCD-GS and representative 3DGS-based methods on real-world scenes. The first column presents the complete scene views, with red boxes indicating the enlarged regions. The remaining columns show the corresponding enlarged results produced by the compared methods. (a) *bicycle*, containing thin spokes and fine contours; (b) *garden*, containing dense vegetation and complex boundaries; (c) *Dr Johnson*, containing door frames, wall boundaries, and local geometric details; and (d) *truck*, containing vehicle contours, window edges, and high-frequency appearance variations

As shown in Figure 3, the original 3DGS tends to produce softened boundaries and blurred fine structures in challenging regions. Its gradient-driven densification

strategy increases the representation capacity of under-optimized regions, but it does not explicitly distinguish texture-dominant regions from structure-dominant regions. Consequently, thin structures and regular boundaries may remain incomplete or over-smoothed.

AbsGS and Pixel-GS recover additional details in some high-frequency regions by improving gradient statistics or introducing pixel-aware density control. However, local noise, broken contours, and unstable boundaries can still occur when dense textures and complex structures appear simultaneously. Perceptual-GS provides clearer results in visually salient regions, but thin structures and regular boundaries may remain insufficiently reconstructed because structural cues are not explicitly incorporated into the complete densification process.

PSCD-GS produces more consistent visual results across the selected examples. In the *bicycle* scene, the wheel spokes and surrounding contours are reconstructed with fewer merged or interrupted structures. In *garden*, vegetation textures are retained more clearly, with less noticeable over-smoothing. In *Dr Johnson*, the door frames and wall boundaries appear sharper and more continuous. In *truck*, the vehicle contours, window edges, and transitions between adjacent body regions are reconstructed with fewer local artifacts.

These qualitative results further support the quantitative improvements reported in Table 2. In particular, the visual comparisons indicate that the gains in SSIM, PSNR, and LPIPS are accompanied by improved preservation of local textures, thin structures, and object boundaries. Meanwhile, PSCD-GS maintains a controlled representation size rather than relying on unrestricted growth in the number of Gaussian primitives.

4.4 Ablation experiments

To analyze the contribution of each component in PSCD-GS, we performed ablation experiments on Mip-NeRF 360, Deep Blending, and Tanks & Temples. In addition to removing the two evidence branches, we further evaluated the continuous role representation and the subsequent densification decisions. This analysis is necessary because the proposed perception-structure role is used not only to characterize Gaussian-level evidence, but also to determine where and how densification is performed.

Specifically, we constructed six ablated variants. The first two variants removed the perceptual evidence branch and the structural evidence branch, respectively. The third variant removed the soft role assignment and replaced the continuous perception-structure role representation with the equal-weight fusion $\tilde{S}_i = (P_i + Q_i)/2$. The remaining three variants independently disabled joint candidate selection, region-adaptive clone/split selection, and role-guided split placement. These components correspond to three successive decisions in the densification process: which Gaussians should be densified, whether clone or split should be applied, and where the split children should be placed.

For all ablations, only the specified component was changed, while the remaining training settings and hyperparameters were kept identical to those of the full model. In the variant without perceptual evidence, the perceptual response and its associated directional contribution were disabled, while the structural branch remained

unchanged. Conversely, the variant without structural evidence retained only the perceptual branch. In the variant without soft role assignment, the equal-weight fused evidence $\tilde{S}_i$ was used in place of the continuous role memberships in the subsequent densification decisions. The variant without joint candidate selection reverted to the original gradient-based candidate criterion of 3DGS. The variant without region-adaptive clone/split selection used the original scale-based operation rule, whereas the variant without role-guided split placement restored the original 3DGS split initialization. No other component was modified in any ablated variant.

Table 3 reports the quantitative results of these variants and the complete PSCD-GS model.

Table 3 Ablation study of PSCD-GS on three benchmarks.

Variant	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$
<i>Mip-NeRF 360</i>			
w/o perceptual evidence	0.8298	27.9476	0.1951
w/o structural evidence	0.8311	28.0283	0.1904
w/o soft role assignment	0.8330	<u>28.1712</u>	0.1882
w/o joint candidate selection	0.8285	27.9028	0.1936
w/o region-adaptive clone/split	0.8366	28.1177	<u>0.1849</u>
w/o role-guided split placement	<u>0.8370</u>	28.1594	0.1864
Full model	0.8396	28.2439	0.1814
<i>Deep Blending</i>			
w/o perceptual evidence	0.9107	29.9657	0.2366
w/o structural evidence	0.9112	30.1158	0.2307
w/o soft role assignment	0.9128	30.1983	<u>0.2277</u>
w/o joint candidate selection	0.9105	30.0474	0.2351
w/o region-adaptive clone/split	0.9148	<u>30.2064</u>	0.2303
w/o role-guided split placement	<u>0.9151</u>	30.1562	0.2289
Full model	0.9184	30.2961	0.2243
<i>Tanks & Temples</i>			
w/o perceptual evidence	0.8739	24.4316	0.1495
w/o structural evidence	0.8702	24.3184	0.1550
w/o soft role assignment	0.8739	<u>24.5361</u>	0.1477
w/o joint candidate selection	0.8726	24.3495	0.1512
w/o region-adaptive clone/split	0.8751	24.5263	<u>0.1438</u>
w/o role-guided split placement	<u>0.8775</u>	24.4982	0.1447
Full model	0.8815	24.5968	0.1407

Note: “w/o” denotes “without”. The best results are shown in bold, and the second-best results are underlined. SSIM, structural similarity index measure; PSNR, peak signal-to-noise ratio; LPIPS, learned perceptual image patch similarity. The upward arrow indicates that a higher value is better, whereas the downward arrow indicates that a lower value is better.

In Mip-NeRF 360, the removal of joint candidate selection results in the largest degradation in SSIM and PSNR, indicating that the identification of reliable and informative densification candidates is particularly important in complex unbounded scenes. In contrast, removing the perceptual evidence branch causes the most pronounced increase in LPIPS. This result suggests that perceptual responses are especially useful for locating visually sensitive regions, such as textured surfaces and

high-frequency appearance variations, which may not be sufficiently characterized by gradient magnitude or structural cues alone. The structural branch also contributes consistently, although its influence is smaller than that of the perceptual branch and the selection of joint candidates on this benchmark.

A similar tendency can be observed on Deep Blending. The perceptual evidence branch and joint candidate selection produce the two largest overall contributions. In particular, removing the perceptual branch leads to a clear reduction in PSNR and a noticeable deterioration in LPIPS, whereas removing joint candidate selection causes the largest decrease in SSIM. These results indicate that appearance-sensitive evidence construction and reliable candidate screening play complementary roles: the former improves the representation of perceptually important details, while the latter prevents densification from being allocated to redundant or low-value primitives.

The behavior on Tanks & Temples is different. Removing the structural evidence branch produces the largest degradation in SSIM, PSNR, and LPIPS. This benchmark contains prominent object boundaries, depth discontinuities, geometric edges, and thin structures, making structural information more influential than on the other two datasets. Joint candidate selection produces the second-largest overall degradation, further confirming that reliable candidate screening remains important even when structural cues dominate. The dataset-dependent variation also indicates that perceptual and structural evidence should not be treated as interchangeable signals.

Removing soft role assignment consistently reduces the quality of reconstruction on all three benchmarks. Although its effect is generally smaller than that of the dominant importance branch or joint candidate selection, the degradation confirms that a continuous perception–structure role is more effective than directly combining the two evidence scores with a fixed fusion rule. The proposed role representation allows each Gaussian to occupy a continuous position in the joint perception–structure space, thereby distinguishing perception-dominant, structure-dominant, and jointly important regions. This provides a more adaptive basis for subsequent densification decisions. The region-adaptive clone/split strategy and role-guided split placement also improve performance, although their contributions are relatively smaller and vary across metrics. This behavior is reasonable because these components operate after the important regions and candidate Gaussians have already been identified. Region-adaptive operation selection determines whether a Gaussian should be cloned or split according to its regional role, while role-guided placement controls the spatial distribution of the generated child Gaussians. Their removal does not eliminate the upstream evidence construction and candidate-selection mechanisms, but it reduces the precision with which new primitives are allocated within local regions.

In general, the ablation results support the contribution of collaboration in the estimation of perceptual evidence, the estimation of structural evidence, the modeling of soft roles, and the role-aware densification decisions. The two evidence branches provide complementary cues, the soft role assignment organizes these cues into a continuous regional representation, and the three densification components progressively determine where to densify, how to densify, and where to place the generated Gaussians. Together, these components extend the original gradient-driven densification mechanism of 3DGS into a perception–structure role-aware decision process.

Detailed per-scene ablation results for all evaluated variants are provided in Online Resource 2.

4.5 Discussion

The experimental results suggest that the improvement of PSCD-GS is associated with explicit perception–structure role modeling rather than with unrestricted primitive growth. The original 3DGS densification strategy uses view-space gradients to identify under-optimized regions, which is effective but does not describe the regional attributes of different Gaussians. In contrast, PSCD-GS assigns each Gaussian a continuous role in the joint perception–structure space and uses this role to guide candidate selection, clone/split decisions, and split placement. This design allows the densification process to respond differently to texture-rich regions, structural boundaries, thin structures, and low-response areas.

The results also show that perceptual and structural signals are complementary. Perceptual evidence primarily reflects texture details, local appearance variations, and visually sensitive regions, whereas structural evidence provides complementary cues for boundaries, contours, thin structures, and geometric transitions. A single cue is therefore insufficient for complex real-world scenes. The proposed soft role assignment avoids treating these cues as fixed-weight scores and instead adapts their contributions according to the regional role of each Gaussian.

PSCD-GS should also be distinguished from compact or efficiency-oriented 3DGS methods. Some compact methods reduce the number of Gaussians through pruning, sampling control, or compression, whereas PSCD-GS focuses on quality-oriented Gaussian allocation during densification. Its goal is not to obtain the minimum representation scale, but to assign representation capacity to regions where additional Gaussians are most beneficial. Therefore, PSCD-GS is complementary to compact Gaussian representations and can be further combined with pruning or compression strategies to improve the trade-off between quality, scale, and efficiency.

There are still several limitations. First, estimating perceptual and structural responses introduces additional training-time computation. Although the perceptual maps are precomputed, the structural responses, multi-view reliability, and contribution-weighted tensor statistics are updated during densification events. These operations do not alter the standard Gaussian representation or introduce an additional neural inference module at rendering time, but their training overhead remains an aspect that could be examined more systematically in future work.

Second, the reliability of structural evidence depends on view coverage, depth quality, and reprojection consistency. Sparse views, reflective surfaces, motion, or noisy boundaries may weaken the cross-view gate and reduce the benefit of role-aware placement. Third, the current framework is evaluated mainly on static real-world scenes; dynamic scenes, online reconstruction, and large-scale environments remain untested. Finally, the role assignment and densification policy still rely on fixed hyperparameters. Adaptive or learning-based policies may improve robustness across scenes and training stages, but such extensions require separate validation.

5 Conclusion

We presented PSCD-GS, a perception–structure collaborative densification framework for 3D Gaussian Splatting. The method maps image-space perceptual responses and multi-view structural evidence to Gaussian primitives, represents their joint state through continuous role memberships, and uses this representation to coordinate candidate selection, clone-or-split decisions, and split placement. In this way, densification is treated as a sequence of related decisions rather than only as gradient-thresholded primitive growth.

Experiments on Mip-NeRF 360, Deep Blending, and Tanks & Temples show that PSCD-GS achieves competitive dataset-level rendering quality with a controlled number of Gaussian primitives. The ablation results support the contributions of the perceptual evidence branch, the structural evidence branch, continuous role assignment, and role-aware densification decisions. Qualitative comparisons further suggest that the method better preserves local textures, object boundaries, and thin structures in challenging scenes. Future work will focus on reducing the additional evidence-construction overhead, extending the method to dynamic and large-scale scenes, and replacing fixed densification hyperparameters with more adaptive policies.

Supplementary information. Online Resource 1 contains the complete per-scene quantitative results for all evaluated methods on Mip-NeRF 360, Deep Blending, and Tanks & Temples.

Online Resource 2 contains the complete per-scene ablation results for all evaluated variants on Mip-NeRF 360, Deep Blending, and Tanks & Temples.

Statements and Declarations

Funding. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests. The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate. Not applicable. This study did not involve human participants or animals.

Consent for publication. Not applicable.

Data Availability. The datasets analyzed during the current study, including Mip-NeRF 360, Deep Blending, and Tanks & Temples, are publicly available from their original repositories and are cited in the manuscript. No new dataset was generated in this study.

Materials Availability. Not applicable.

Code Availability. The source code developed for this study is not publicly available at the time of submission. It will be released in a public repository upon acceptance. During peer review, the code is available from the corresponding author upon reasonable request.

Author Contributions. Deyong Shang conceived the study, developed the methodology and software, performed the experiments, analyzed the results, prepared the figures, and wrote the original manuscript. Jiaxin Shi contributed to data preparation, experimental validation, result analysis, and manuscript revision. Lianyu Mu contributed to data preparation, result verification, visualization, and manuscript revision. All authors read and approved the final manuscript.

References

- [1] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM* **65**(1), 99–106 (2021) <https://doi.org/10.1145/3503250>
- [2] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42**(4), 139 (2023) <https://doi.org/10.1145/3592433>
- [3] Ye, Z., Li, W., Liu, S., Qiao, P., Dou, Y.: AbsGS: Recovering fine details in 3D gaussian splatting. In: *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 1053–1061. Association for Computing Machinery, Melbourne, VIC, Australia (2024). <https://doi.org/10.1145/3664647.3681361>
- [4] Zhang, Z., Hu, W., Lao, Y., He, T., Zhao, H.: Pixel-GS: Density control with pixel-aware gradient for 3D gaussian splatting. In: *Computer Vision – ECCV 2024*, pp. 326–342. Springer, Milan, Italy (2024)
- [5] Fang, G., Wang, B.: Mini-Splatting: Representing scenes with a constrained number of gaussians. In: *Computer Vision – ECCV 2024*, pp. 165–181. Springer, Milan, Italy (2024)
- [6] Mallick, S.S., Goel, R., Kerbl, B., Steinberger, M., Carrasco, F.V., Torre, F.: Taming 3DGS: High-quality radiance fields with limited resources. In: *SIGGRAPH Asia 2024 Conference Papers*, pp. 1–11. Association for Computing Machinery, Tokyo, Japan (2024). <https://doi.org/10.1145/3680528.3687694>
- [7] Kheradmand, S., Rebain, D., Sharma, G., Sun, W., Tseng, Y.-C., Isack, H., Kar, A., Tagliasacchi, A., Yi, K.M.: 3D gaussian splatting as markov chain monte carlo. *Advances in Neural Information Processing Systems* **37**, 80965–80986 (2024)
- [8] Wang, P., Wang, Y., Wang, D., Mohan, S., Fan, Z., Wu, L., Cai, R., Yeh, Y.-Y., Wang, Z., Liu, Q., Ranjan, R.: Steepest descent density control for compact 3D gaussian splatting. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26663–26672 (2025)
- [9] Zhou, H., Ni, Z.: Perceptual-GS: Scene-adaptive perceptual densification for Gaussian splatting. In: *Proceedings of the 42nd International Conference on*

Machine Learning. Proceedings of Machine Learning Research, vol. 267, pp. 79089–79114. PMLR, Vancouver, BC, Canada (2025)

- [10] Deng, X., Diao, C., Li, M., Yu, R., Xu, D.: Improving densification in 3D gaussian splatting for high-fidelity rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Findings, pp. 223–232 (2026)
- [11] 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 (2022)
- [12] Hedman, P., Philip, J., Price, T., Frahm, J.-M., Drettakis, G., Brostow, G.: Deep blending for free-viewpoint image-based rendering. *ACM Transactions on Graphics* **37**(6), 257 (2018) <https://doi.org/10.1145/3272127.3275084>
- [13] Knapitsch, A., Park, J., Zhou, Q.-Y., Koltun, V.: Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics* **36**(4), 78 (2017) <https://doi.org/10.1145/3072959.3073599>
- [14] Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-Splatting: Alias-free 3D gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 19447–19456 (2024)
- [15] Cheng, K., Long, X., Yang, K., Yao, Y., Yin, W., Ma, Y., Wang, W., Chen, X.: GaussianPro: 3D gaussian splatting with progressive propagation. In: Proceedings of the 41st International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 235, pp. 8123–8140. PMLR, Vienna, Austria (2024)
- [16] Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., Dai, B.: Scaffold-GS: Structured 3D gaussians for view-adaptive rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 20654–20664 (2024)
- [17] 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), pp. 586–595 (2018)
- [18] Dong, Z., Jiang, J., Chen, Y., Zhang, J., Jiang, K., Liu, X.: Reframing gaussian splatting densification with complexity-density consistency of primitives. In: Advances in Neural Information Processing Systems, vol. 38 (2025)
- [19] Guédon, A., Lepetit, V.: SuGaR: Surface-aligned gaussian splatting for efficient 3D mesh reconstruction and high-quality mesh rendering. In: Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5354–5363 (2024)

- [20] Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2D gaussian splatting for geometrically accurate radiance fields. In: SIGGRAPH 2024 Conference Papers, pp. 1–11. Association for Computing Machinery, Denver, CO, USA (2024). <https://doi.org/10.1145/3641519.3657428>
- [21] Yu, Z., Sattler, T., Geiger, A.: Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Transactions on Graphics* **43**(6), 271 (2024) <https://doi.org/10.1145/3687937>
- [22] Lyu, L., Tewari, A., Chen, J., Leimkühler, T., Theobalt, C.: Faster 3D gaussian splatting convergence via structure-aware densification. In: SIGGRAPH 2026 Conference Papers, pp. 1–10. Association for Computing Machinery, Los Angeles, CA, USA (2026). <https://doi.org/10.1145/3799902.3811212>

Author biographies



Deyong Shang is a professor and doctoral supervisor in the Department of Mechatronic Engineering, School of Mechanical and Electrical Engineering, China University of Mining and Technology-Beijing. He received his Ph.D. in Mechanical Design and Theory from China University of Mining and Technology-Beijing in 2016. His research interests include robotics and its applications, intelligent mining equipment, image processing and recognition, modern design methods, and safety technologies for coal-mining machinery. He has led and participated in national, provincial, and industry-funded research projects.



Jiaxin Shi is a master's student in the School of Mechanical and Electrical Engineering at China University of Mining and Technology-Beijing. Her research interests include computer vision, 3D Gaussian Splatting, and simultaneous localization and mapping. Her current research focuses on three-dimensional scene representation, neural rendering, visual perception, and high-fidelity scene reconstruction in complex environments.



Lianyu Mu is a master's student in the School of Mechanical and Electrical Engineering at China University of Mining and Technology-Beijing. Her research interests include computer vision, path planning, and intelligent robotic systems. Her current research focuses on image-based scene understanding, visual perception, and efficient planning methods for autonomous navigation and robotic applications.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.zip](#)