Geometric Probing for Isotropic Optimization Manifold in Sparse-View 3D Gaussian Splatting

Yunlong Zhao[1], Xiaoheng Deng[1]*, Hongyan Xu[2], Yichao Cao[1], Keke Huang[1], Shuo Yang[3], Xiangjian He[4], Lei Fan[5], Zhuohua Qiu[6], and Xiu Su[1]†

[1] Central South University, Changsha, China
[2] Tianjin University, Tianjin, China
[3] Harbin Institute of Technology (Shenzhen), Shenzhen, China
[4] University of Nottingham Ningbo China, Ningbo, China
[5] UNSW Sydney, Sydney, Australia
[6] Shenzhen University, Shenzhen, China

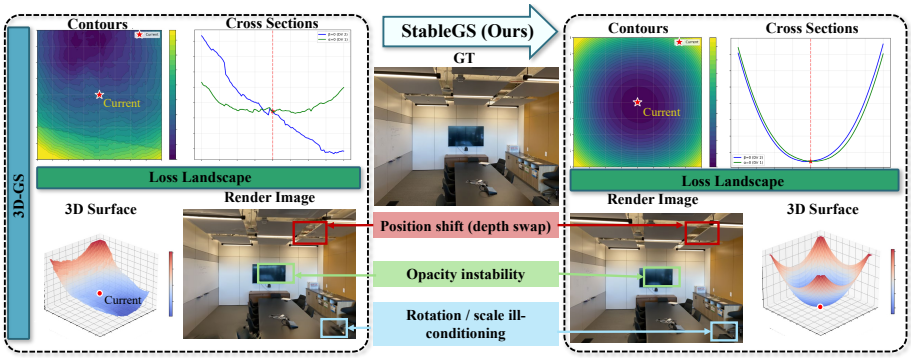

Fig. 1: Optimization manifold isotropy correlates with rendering stability. Left (Anisotropy): 3DGS exhibits directionally anisotropic loss surfaces—elongated contours and sharp cross-sectional curvature manifest as unstable geometry and rendering artifacts. Diagnostic probing identifies three instability sources: position-induced depth swaps, opacity variations, and rotation/scale ill-conditioning. **Right (Isotropy):** StableGS regularizes toward an isotropic manifold with circular contours and symmetric cross-sections, yielding stable geometry and artifact-free rendering.

Abstract. 3D Gaussian Splatting (3DGS) achieves impressive rendering quality but degrades drastically under sparse views. Through optimization manifold analysis, we reveal that sparse supervision produces anisotropic loss surfaces where rendering quality collapses sharply along three geometric fragility axes: position shifts, rotation/scale ill-conditioning, and opacity instability. Guided by this analysis, we propose Stable 3DGS (StableGS), which steers optimization toward isotropic stability via geometric probing. Unlike the generic, uniform smoothing of methods like SAM [8], which cannot fit 3DGS, StableGS uses domain-specific geometric probes to target the splatting operator's unique instabilities. Our approach comprises: (1) **Attribute-Space Probing** that varies each

*Corresponding authors.
†Corresponding authors.

Gaussian's attributes based on Hessian-informed sensitivity analysis to measure rendering fragility, then minimizes loss at probe points; (2) **Viewpoint-Space Probing** that samples geometrically critical poses and enforces consistency between base and probed configurations. StableGS significantly advances state-of-the-art (SOTA) performance across multiple benchmarks, achieving substantial improvements such as +0.64dB PSNR on DTU [14] with 74% anisotropy reduction in optimization manifold. Code is available at `https://github.com/zyl123456aB/StableGS`.

Keywords: Sparse-view novel view synthesis· 3D Gaussian Splatting · Isotropic stability

1 Introduction

Reconstructing high-fidelity 3D scenes from sparse input views remains a fundamental challenge in computer vision, with broad applications [21–23, 45] in autonomous driving [46], augmented reality, and robotics. Acquiring dense multi-view captures is often impractical due to constraints in time, cost, or physical accessibility, making sparse-view novel view synthesis (typically 3-9 images) increasingly critical. Existing methods struggle with artifacts such as floaters and view-dependent inconsistencies under such constraints.

Recent methods address sparse-view degradation through various strategies. NeRF-based approaches [10, 13, 29, 33, 34, 48, 51] employ semantic priors and depth constraints, while 3DGS-based methods [4, 6, 18, 20, 25, 35, 61] leverage external depth supervision or multi-stage training. However, existing methods treat symptoms through task-specific designs (injecting external supervision, applying heuristic regularization [54], or stochastically sampling model variations [60]) without understanding why sparse supervision causes 3DGS to fail.

We investigate this question through optimization manifold analysis, revealing that sparse training produces **directionally anisotropic** loss surfaces where the splatting operator exhibits drastically different sensitivities across parameter directions. By probing through systematic Gaussian attribute variations, we identify three geometric fragility axes where the splatting pipeline becomes critically unstable: First, position shifts along viewing rays cause depth swaps that disrupt occlusion ordering. Second, rotation and scale adjustments trigger projection ill-conditioning. Third, opacity variations destabilize alpha-blending. These instabilities manifest as floaters, splat collapse/explosion, and view-dependent visibility encoding. This directional fragility, where rendering quality remains stable in certain directions but collapses sharply along others, explains why standard optimization converges to solutions that work for training views but fail catastrophically on novel viewpoints.

Our analysis connects to optimization manifold geometry and generalization [8,19,26,56]. Traditional methods like SAM [8] seek flat minima by minimizing worst-case loss within a parameter neighborhood, improving generalization in classification tasks. However, they treat all parameters uniformly and target isotropic flatness in all directions. In contrast, 3DGS exhibits **structured**

anisotropy specific to the splatting operator: depth-direction position shifts cause fundamentally different failures (occlusion flips) than lateral shifts (coverage changes), and these require different probing strategies. Moreover, SAM operates purely in parameter space, whereas 3DGS rendering quality depends critically on geometric relationships (depth ordering, projection conditioning) and viewpoint consistency that cannot be captured by parameter-space probing.

Guided by this analysis, we propose Stable 3DGS (*StableGS*), which uses geometric probes as diagnostic tools to map rendering fragility and then steers optimization toward isotropic stability. Rather than injecting external depth priors [7] or stochastically exploring parameters [60], we systematically probe the optimization manifold to identify rendering-fragile directions and explicitly optimize to avoid these regions. **Attribute-Space Probing** uses Hessian trace estimation to identify sensitive attributes, then perturbs each Gaussian along these vulnerable directions and minimizes worst-case rendering loss. **Viewpoint-Space Probing** samples geometrically critical poses (occlusion boundaries, parallax discontinuities) and enforces rendering consistency between base and probed configurations, preventing convergence to view-specific sharp minima. Together, ASP and VSP reshape the optimization manifold from directionally anisotropic to isotropic, yielding intrinsically stable splatting operators. The key contributions of this work are as follows:

- We identify three geometric fragility axes in 3DGS under sparse supervision and propose StableGS, which steers optimization toward isotropic stability without external supervision;
- We introduce Attribute-Space Probing (ASP), which perturbs Gaussians along Hessian-identified vulnerable directions and minimizes worst-case rendering loss, achieving 74% anisotropy reduction;
- We introduce Viewpoint-Space Probing (VSP), which enforces rendering consistency at geometrically critical poses, preventing convergence to view-specific sharp minima;
- We improve PSNR over the state-of-the-art by 0.33 dB, 0.64 dB, and 0.25 dB on the LLFF, DTU, and Mip-NeRF360 datasets, respectively.

2 Related Work

2.1 Sparse-View Novel View Synthesis

Neural Radiance Fields [31] severely overfit under sparse inputs. NeRF-based methods address this through external supervision: PixelNeRF [55] uses pretrained encoders, DietNeRF [13] leverages CLIP [38] semantics, SparseNeRF [48] distills depth ranking from monocular estimators [40], RegNeRF [34] applies patch regularization, and FreeNeRF [54] employs frequency regularization.

3D Gaussian Splatting [16] also degrades under sparse views. Depth-based methods like DNGaussian [20] and FSGS [61] leverage monocular depth estimators [39, 40] but depend on error-prone priors [41, 52]. CoR-GS [57] co-regularizes dual fields through registration and rendering agreement. Dropout

methods [36, 53] randomly eliminate Gaussians inspired by DropConnect [47]. However, these methods lack systematic analysis of *why* sparse views cause splatting failures. Unlike prior works addressing symptoms, we identify three failure modes—depth ordering instability, projection degeneracy, and alpha-blending incoherence—and stabilize the differentiable splatting operator.

2.2 Rendering Stability and Robustness

Robustness has been studied in adversarial training [9, 32] and consistency regularization [24, 50]. In neural rendering, recent work explores input-level stability: NeRF in the Dark [27] handles noisy images, Self-Calibrating NeRF [15] addresses pose errors, Gaussian Activated NeRF [6] studies activation choices, and Scene Representation Transformer [42] uses set-latent representations for sparse-view robustness. Hash-based methods [11, 33] analyze encoding stability, while uncertainty quantification [43, 44] estimates prediction confidence. However, these focus on input perturbations or encoding rather than parameter-space stability of the rendering operator.

Unlike general optimization [8, 59] seeking flat minima, we directly stabilize the splatting operator through probing-aware training and cross-view consistency, addressing the cause of sparse-view degradation.

Algorithm 1: StableGS Training Pipeline

Require: Images $\{\mathbf{I}_m\}$, poses $\{\mathbf{P}_m\}$
Ensure: Optimized Gaussians $\boldsymbol{\Theta}^*$
Initialize $\mathcal{G}$ from SfM
Warm-up (1–1000):
for $t = 1$ to 1000 **do**
  Sample m, render $\hat{\mathbf{I}}_m$
  Update $\mathcal{G}$ with $\nabla\mathcal{L}_{\text{recon}}$
end for
Probing (1001–$t_{\max}$):
for $t = 1001$ to $t_{\max}$ **do**
  Sample m, render $\hat{\mathbf{I}}_m$
  Update τ_ϕ (every 100 iters)
  Compute ρ_ϕ (Eq. (5))
  Perturb: $\tilde{\boldsymbol{\Theta}} \leftarrow$ Eqs. (8)–(13)
  Render $\tilde{\mathbf{I}}_m$, compute $\mathcal{L}_{\text{ASP}}$
  Sample K cameras $\{\mathbf{P}'_k\}$
  Render $\boldsymbol{\Theta}, \tilde{\boldsymbol{\Theta}}$, compute $\mathcal{L}_{\text{VSP}}$
  $\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda_{\text{A}}\mathcal{L}_{\text{ASP}} + \lambda_{\text{V}}\mathcal{L}_{\text{VSP}}$
  Update $\mathcal{G}$ with $\nabla\mathcal{L}$
  Densify/prune (every 100 iters)
end for

Second-order optimization. Recent works accelerate 3DGS with second-order updates: 3DGS^2 [17] and matrix-free Levenberg–Marquardt with residual sampling [37] approximate Newton/LM steps for faster, denser-view fitting. StableGS is *not* a second-order optimizer: it never forms or inverts the Hessian and never solves $\mathbf{H}^{-1}\nabla\mathcal{L}$; instead it uses cheap Hessian trace/eigenvalue *diagnostics* (Eqs. (3), (6)) to place geometric probes for sparse-view regularization. The two are orthogonal—second-order solvers ask *how fast* 3DGS converges, whereas StableGS asks *which* sparse-view minimum it converges to. Moreover, since ASP and VSP operate purely through differentiable rendering losses and parameter-space probes without modifying the rasterizer, StableGS is complementary to renderer-side advances such as Mip-Splatting and 2D-GS: any can serve as the back-end while our probes reshape the manifold, so the contributions stack rather than substitute.

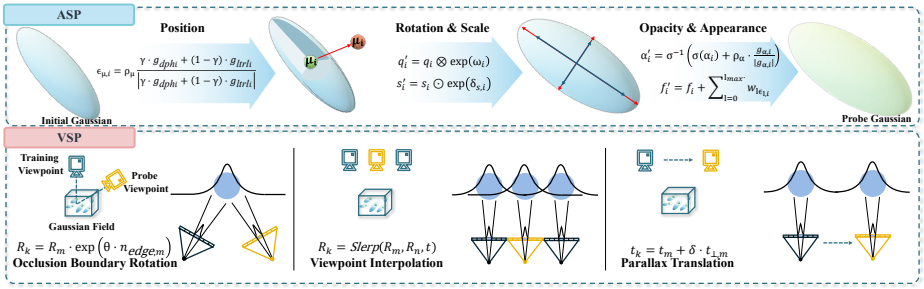

Fig. 2: Overview of ASP and VSP. ASP and VSP steer 3DGS toward isotropically stable optimization manifold for robust novel view synthesis.

3 Method

As shown in Figure 2, we present StableGS with ASP and VSP to address sparse-view 3DGS degradation through geometric probing. Our approach: (1) analyzes optimization manifold anisotropy to identify geometric fragility axes, (2) uses Hessian-informed probes as diagnostic tools to map rendering-fragile regions, and (3) steers optimization toward isotropic stability.

3.1 Preliminaries: 3D Gaussian Splatting

Following 3DGS [16], we represent a scene using N anisotropic Gaussians, where each Gaussian i is parameterized by: position $\boldsymbol{\mu}_i \in \mathbb{R}^3$, scale $\mathbf{s}_i \in \mathbb{R}^3$, rotation quaternion $\mathbf{q}_i \in \mathbb{R}^4$ (unit quaternion), opacity $\alpha_i \in \mathbb{R}$ (logit space), and spherical harmonic coefficients $\mathbf{f}_i \in \mathbb{R}^K$ for view-dependent appearance.

We denote all parameters as $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_i, \mathbf{s}_i, \mathbf{q}_i, \mathbf{f}_i, \alpha_i\}_{i=1}^N$ and training camera poses as $\{\mathbf{P}_m\}_{m=1}^M$, where $\mathbf{P}_m = \{\mathbf{R}_m, \mathbf{t}_m\}$ with $\mathbf{R}_m \in SO(3)$ denoting the **World-to-Camera rotation matrix** and $\mathbf{t}_m \in \mathbb{R}^3$ denoting the **camera center in world coordinates**.

The 3D covariance matrix is $\boldsymbol{\Sigma}_i = \mathbf{R}(\mathbf{q}_i)\mathbf{S}(\mathbf{s}_i)\mathbf{S}(\mathbf{s}_i)^T\mathbf{R}(\mathbf{q}_i)^T$. To render pixel $\mathbf{p}$ from camera $\mathbf{P}_m$, the splatting pipeline: (1) projects 3D covariance to 2D via $\boldsymbol{\Sigma}_i' = \mathbf{J}_i\mathbf{W}_m\boldsymbol{\Sigma}_i\mathbf{W}_m^T\mathbf{J}_i^T$ (where $\mathbf{W}_m$ is the world-to-camera rotation and $\mathbf{J}_i$ the projective Jacobian at $\boldsymbol{\mu}_i$ [62]); (2) depth-sorts Gaussians by their camera-space depth $d_i = (\mathbf{R}_m(\boldsymbol{\mu}_i - \mathbf{t}_m))_z$ (i.e., the z-coordinate of the Gaussian center in camera coordinates); (3) alpha-composites colors:

$$\mathbf{C}(\mathbf{p}; \boldsymbol{\Theta}, \mathbf{P}_m) = \sum_{i \in \mathcal{N}(\mathbf{p})} \mathbf{c}_i(\mathbf{d}_m) \cdot \tilde{\alpha}_i \prod_{j=1}^{i-1}(1 - \tilde{\alpha}_j), \tag{1}$$

where $\mathcal{N}(\mathbf{p})$ denotes the **depth-sorted ordered list** of Gaussians whose 2D footprints cover pixel $\mathbf{p}$, $\mathbf{c}_i(\mathbf{d}_m)$ decodes color from SH coefficients, and $\tilde{\alpha}_i = \sigma(\alpha_i)\mathcal{G}_i^{2\mathrm{D}}(\mathbf{p})$ combines opacity with 2D Gaussian footprint.

Standard training minimizes photometric loss:

$$\mathcal{L}_{\text{recon}}(\boldsymbol{\Theta}) = \frac{1}{M}\sum_{m=1}^{M}\left[\ell_1(\hat{\mathbf{I}}_m, \mathbf{I}_m) + \lambda_{\text{SSIM}}\mathcal{L}_{\text{SSIM}}(\hat{\mathbf{I}}_m, \mathbf{I}_m)\right]. \tag{2}$$

3.2 Optimization Manifold Diagnosis

We characterize the optimization manifold's directional stability through local curvature. For parameter configuration $\boldsymbol{\Theta}$, the Hessian matrix $\mathbf{H} = \nabla^2_{\boldsymbol{\Theta}}\mathcal{L}(\boldsymbol{\Theta})$ captures second-order sensitivity. We define the **anisotropy index**:

$$\mathcal{A}(\boldsymbol{\Theta}) = \frac{\lambda_{\max}(\mathbf{H})}{\lambda_{\min}(\mathbf{H})}, \tag{3}$$

measuring the condition number of the loss surface. High $\mathcal{A}$ indicates directional fragility: large $\lambda_{\max}$ implies steep curvature along certain eigenvectors, while small $\lambda_{\min}$ implies flat, low-curvature directions elsewhere—this coexistence of steep and flat directions is the defining signature of directional anisotropy. We approximate $\lambda_{\max}$ and $\lambda_{\min}$ via Power Iteration [28], reusing the Hessian-vector product (HvP) operator established in Eq. (6).

For each attribute type $\phi \in \{\mu, q, s, \alpha, f\}$, we define the attribute-specific Hessian sub-matrix $\mathbf{H}_\phi = \nabla^2_{\boldsymbol{\theta}^{(\phi)}}\mathcal{L}$, where $\boldsymbol{\theta}^{(\phi)}$ collects all parameters of type ϕ across all Gaussians. These sub-matrices capture the second-order sensitivity of the loss to each attribute type independently.

Position shift (depth swap). During the depth-sort step, Gaussians near occlusion boundaries exhibit unstable rankings. The depth $d_i = (\mathbf{R}_m(\boldsymbol{\mu}_i - \mathbf{t}_m))_z$ is the scalar z-coordinate of Gaussian i in the camera frame, where $\mathbf{R}_m$ rotates from world to camera coordinates and $\mathbf{t}_m$ is the camera center. Consider a position probe $\boldsymbol{\mu}_i \to \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_\mu$ with $\|\boldsymbol{\epsilon}_\mu\| \sim 0.1$ world units. If $\boldsymbol{\epsilon}_\mu$ aligns with the viewing ray $\mathbf{v}_{\text{cam}} = \boldsymbol{\mu}_i - \mathbf{t}_m$, the depth changes as:

$$d_i \to d_i + \Delta d, \quad \Delta d = (\mathbf{R}_m \boldsymbol{\epsilon}_\mu)_z, \tag{4}$$

where Δd is the projection of the perturbation onto the camera's optical axis. When $|\Delta d|$ exceeds the inter-Gaussian depth gap, Gaussian i swaps its rank in the ordered list $\mathcal{N}(\mathbf{p})$, causing it to abruptly appear in front of or behind nearby surfaces—manifesting as floaters. The gradient magnitude along the depth direction $\|\nabla_{\boldsymbol{\mu}_i}\mathcal{L} \cdot \hat{\mathbf{v}}_{\text{cam}}\|$ (where $\hat{\mathbf{v}}_{\text{cam}} = \mathbf{v}_{\text{cam}}/\|\mathbf{v}_{\text{cam}}\|$) is typically $100\times$ larger than lateral directions, contributing to large eigenvalues in $\mathbf{H}_\mu$.

Rotation / Scale Ill-conditioning. The projected covariance $\boldsymbol{\Sigma}'_i$ becomes ill-conditioned when $\boldsymbol{\Sigma}_i$ has extreme aspect ratios or aligns adversarially with the camera. Small rotation/scale probes can cause eigenvalue collapse: $\lambda_{\min}(\boldsymbol{\Sigma}'_i) \to 0$ or $\lambda_{\max}(\boldsymbol{\Sigma}'_i) \to \infty$, collapsing splats into lines or exploding them to cover the screen. We quantify this via the condition number: $\kappa(\boldsymbol{\Sigma}'_i) = \frac{\lambda_{\max}(\boldsymbol{\Sigma}'_i)}{\lambda_{\min}(\boldsymbol{\Sigma}'_i)}$. Gaussians with $\kappa(\boldsymbol{\Sigma}'_i) > 100$ dominate the high-curvature eigenspace of $\mathbf{H}_q$ and $\mathbf{H}_s$, creating sharp loss ridges.

Opacity instability. Under sparse supervision, opacity α_i often encodes view-specific visibility. A Gaussian may adopt $\sigma(\alpha_i) \approx 1$ to block certain training views while $\sigma(\alpha_i) \approx 0$ elsewhere. This produces inconsistent transmittance products $T_i = \prod_{j=1}^{i-1}(1 - \tilde{\alpha}_j)$ on novel views, causing ghosting artifacts. High curvature in $\mathbf{H}_\alpha$ indicates opacity-related fragility.

These fragility axes share a common cause: the loss surface $\mathcal{L}(\boldsymbol{\Theta})$ exhibits steep gradients along fragile directions coexisting with flat regions elsewhere—the directional anisotropy characterized by high $\mathcal{A}(\boldsymbol{\Theta})$.

3.3 Attribute-Space Probing (ASP)

Attribute Probing Degree. ASP diagnoses rendering fragility by perturbing Gaussian attributes in directions informed by local curvature. For each attribute type $\phi \in \{\mu, q, s, \alpha, f\}$ (position, rotation, scale, opacity, appearance), we compute the attribute curvature scale: $\tau_\phi = \sqrt{\mathrm{tr}(\mathbf{H}_\phi)/d_\phi}$, where $\mathbf{H}_\phi$ is the Hessian sub-matrix for attribute ϕ and d_ϕ is its dimensionality (e.g., $d_\mu = 3$, $d_q = 4$, $d_\alpha = 1$). This root-mean-squared eigenvalue measures average curvature magnitude along the ϕ-manifold. The probe magnitude is set inversely proportional to curvature:

$$\rho_\phi = \frac{c_\phi}{\tau_\phi + \epsilon}, \tag{5}$$

where c_ϕ is a global scaling constant (shared across all Gaussians for attribute ϕ) and $\epsilon = 10^{-6}$ prevents division by zero. In steep regions (large τ_ϕ), small ρ_ϕ prevents overshooting sharp gradients; in flat regions (small τ_ϕ), large ρ_ϕ enables effective exploration.

We approximate $\mathrm{tr}(\mathbf{H}_\phi)$ via Hutchinson's estimator [12]:

$$\mathrm{tr}(\mathbf{H}_\phi) \approx \frac{1}{B}\sum_{b=1}^{B} \mathbf{z}_b^T \mathbf{H}_\phi \mathbf{z}_b, \quad \mathbf{z}_b \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d_\phi}), \tag{6}$$

where B random probes yield relative variance in the trace estimate. The Hessian-vector product $\mathbf{H}_\phi \mathbf{z}_b = \nabla_{\boldsymbol{\theta}^{(\phi)}}(\nabla_{\boldsymbol{\theta}^{(\phi)}}\mathcal{L} \cdot \mathbf{z}_b)$ is computed via automatic differentiation without explicit materialization of $\mathbf{H}_\phi$. The same HvP operator is reused in Power Iteration to estimate $\lambda_{\max}(\mathbf{H})$ and $\lambda_{\min}(\mathbf{H})$ for the anisotropy $\mathcal{A}$.

Position Probing Direction. To diagnose depth ordering stability, we decompose the position gradient into depth and lateral components. Let $\mathbf{v}_{\mathrm{cam},i} = \boldsymbol{\mu}_i - \mathbf{t}_m$ denote the (unnormalized) ray from camera center $\mathbf{t}_m$ to Gaussian i, with unit direction $\hat{\mathbf{v}}_{\mathrm{cam},i} = \mathbf{v}_{\mathrm{cam},i}/\|\mathbf{v}_{\mathrm{cam},i}\|$. The depth component is the orthogonal projection of the gradient onto this viewing ray:

$$\mathbf{g}_{\mathrm{depth},i} = \frac{(\nabla_{\boldsymbol{\mu}_i}\mathcal{L}_{\mathrm{recon}})^T \mathbf{v}_{\mathrm{cam},i}}{\|\mathbf{v}_{\mathrm{cam},i}\|^2}\mathbf{v}_{\mathrm{cam},i} = \left((\nabla_{\boldsymbol{\mu}_i}\mathcal{L}_{\mathrm{recon}})^T \hat{\mathbf{v}}_{\mathrm{cam},i}\right)\hat{\mathbf{v}}_{\mathrm{cam},i}. \tag{7}$$

The lateral component is the residual: $\mathbf{g}_{\mathrm{lateral},i} = \nabla_{\boldsymbol{\mu}_i}\mathcal{L}_{\mathrm{recon}} - \mathbf{g}_{\mathrm{depth},i}$.

We construct depth-biased probes using ρ_μ from Eq. (5), with depth-direction weight $\gamma = 0.7$ to specifically stress occlusion boundaries where depth ordering is most fragile:

$$\boldsymbol{\epsilon}_{\mu,i} = \rho_\mu \frac{\gamma \cdot \mathbf{g}_{\mathrm{depth},i} + (1-\gamma)\cdot \mathbf{g}_{\mathrm{lateral},i}}{\|\gamma \cdot \mathbf{g}_{\mathrm{depth},i} + (1-\gamma)\cdot \mathbf{g}_{\mathrm{lateral},i}\|}, \tag{8}$$

where $\gamma \in (0.5, 1)$ emphasizes depth-direction perturbations (which alter sorting rank in $\mathcal{N}(\mathbf{p})$) over lateral perturbations (which only affect pixel coverage). Setting $\gamma = 0.7$ (validated via ablation in Sec. 4.2) produces probes aligned with the dominant failure mode.

Rotation and Scale Probing Direction. For rotation, we sample probes in the tangent space of $SO(3)$. Concretely, for each Gaussian i, we draw a perturbation $\boldsymbol{\omega}_i \in \mathbb{R}^3$ from the Lie algebra $\mathfrak{so}(3)$ and apply it via the exponential map $\exp : \mathfrak{so}(3) \to SO(3)$, then compose with the current rotation quaternion via quaternion multiplication $\otimes$:

$$\mathbf{q}'_i = \mathbf{q}_i \otimes \exp(\boldsymbol{\omega}_i), \quad \boldsymbol{\omega}_i \sim \mathcal{N}(\mathbf{0}, (\rho_{q,i})^2 \mathbf{I}_3), \tag{9}$$

where $\exp(\boldsymbol{\omega}) \in \mathbb{R}^4$ converts the axis-angle vector $\boldsymbol{\omega} \in \mathbb{R}^3$ to a unit quaternion, and $\otimes$ denotes Hamilton quaternion multiplication. The per-Gaussian magnitude $\rho_{q,i}$ is adaptively scaled by the projection condition number:

$$\rho_{q,i} = \rho_q \cdot \min\left(\frac{\kappa(\boldsymbol{\Sigma}'_i) - 1}{100}, 1\right), \tag{10}$$

where ρ_q is the base magnitude from Eq. (5) and $\kappa(\boldsymbol{\Sigma}'_i) = \lambda_{\max}(\boldsymbol{\Sigma}'_i)/\lambda_{\min}(\boldsymbol{\Sigma}'_i)$. When $\boldsymbol{\Sigma}_i$ has extreme aspect ratios or aligns with the viewing direction, $\kappa(\boldsymbol{\Sigma}'_i) \gg 1$ and projection becomes ill-conditioned—small rotation changes cause large 2D footprint variations. Larger probes for high-κ Gaussians force the model to handle these near-degenerate cases.

For scale, we probe in log-space to preserve positivity of scale values:

$$\mathbf{s}'_i = \mathbf{s}_i \odot \exp(\boldsymbol{\delta}_{s,i}), \quad \boldsymbol{\delta}_{s,i} \sim \mathcal{N}(\mathbf{0}, \rho_s^2 \mathbf{I}_3), \tag{11}$$

where $\odot$ denotes element-wise multiplication and $\exp(\cdot)$ is applied element-wise.

Opacity and Appearance Probing Direction. For opacity, we probe directly in logit space using ρ_α from Eq. (5). Since α_i is the logit-space opacity and $g_{\alpha,i} = \partial \mathcal{L}_{\text{recon}}/\partial \alpha_i$ is scalar, the normalized gradient direction reduces to its sign:

$$\alpha'_i = \alpha_i + \rho_\alpha \cdot \frac{g_{\alpha,i}}{\|g_{\alpha,i}\|} = \alpha_i + \rho_\alpha \operatorname{sign}(g_{\alpha,i}). \tag{12}$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid. Probing in logit space guarantees the perturbed opacity $\sigma(\alpha'_i) \in (0, 1)$ by construction, avoiding any evaluation of σ^{-1} at the boundary. For SH coefficients, we apply frequency-weighted probes using ρ_f from Eq. (5):

$$\mathbf{f}'_i = \mathbf{f}_i + \sum_{\ell=0}^{\ell_{\max}} w_\ell \boldsymbol{\epsilon}_\ell, \quad w_\ell = \rho_f \cdot (1 + \ell/\ell_{\max}), \quad \boldsymbol{\epsilon}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{2\ell+1}), \tag{13}$$

where ℓ is the SH degree, $2\ell + 1$ is the number of coefficients in band ℓ, and w_ℓ increases with frequency. Higher-frequency bands ($\ell \geq 2$) are more prone to view-specific memorization; stronger probing ($w_3 = 2w_0$ for degree-3 SH) prevents overfitting.

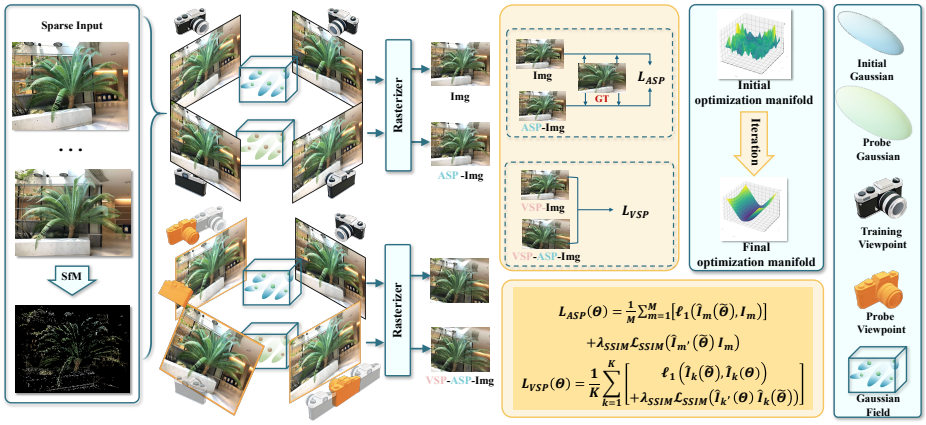

Fig. 3: Overview of StableGS optimization pipeline. Total loss combines reconstruction, attribute fragility, and viewpoint consistency terms.

Given these attribute-specific probes, we perturb all five attributes of all N Gaussians *simultaneously* to form a single global perturbed configuration:

$$\tilde{\boldsymbol{\Theta}} = \{\boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{\mu,i},\ \mathbf{q}_i',\ \mathbf{s}_i',\ \alpha_i',\ \mathbf{f}_i'\}_{i=1}^{N}. \tag{14}$$

This is not per-Gaussian or per-attribute probing; each ASP step therefore costs only *one* additional rasterization rather than $N \times 5$, with per-axis effects isolated only for analysis in Sec. 4.2. We evaluate rendering loss at this probed configuration using the standard 3DGS rasterizer (Eq. (1)):

$$\mathcal{L}_{\text{ASP}}(\boldsymbol{\Theta}) = \frac{1}{M}\sum_{m=1}^{M}\mathcal{L}\left(\hat{\mathbf{I}}_m(\tilde{\boldsymbol{\Theta}}), \mathbf{I}_m\right). \tag{15}$$

Minimizing $\mathcal{L}_{\text{ASP}}$ avoids geometrically fragile configurations.

3.4 Viewpoint-Space Probing (VSP)

While ASP stabilizes rendering on training views, it does not prevent convergence to view-specific sharp minima. Models can memorize training views while occupying narrow, directionally fragile basins that fail on novel viewpoints. We address this through *viewpoint-space probing*—sampling geometrically critical poses to diagnose whether the model generalizes beyond training views.

For each camera $\mathbf{P}_m = \{\mathbf{R}_m, \mathbf{t}_m\}$, we construct a single *composite novel view* $\mathbf{P}_k$ that integrates three geometric probing modes. Rather than applying every probing mode independently, we combine them into one probe per training view, maintaining computational efficiency while testing multiple geometric properties.

(1) Occlusion Boundary Rotation. We identify occlusion boundaries via depth gradients. Let $\hat{\mathbf{D}}_m(\boldsymbol{\Theta})$ be the rendered depth map. We compute the dominant edge normal:

$$\mathbf{n}_{\text{edge},m} = \frac{\nabla\hat{\mathbf{D}}_m}{\|\nabla\hat{\mathbf{D}}_m\|}, \tag{16}$$

which identifies directions perpendicular to occlusion boundaries where depth ordering is most ambiguous.

(2) Viewpoint Interpolation/Extrapolation. To test geometry completeness, we sample a target pose via spherical linear interpolation (Slerp) with a neighboring training camera $\mathbf{P}_n$ (the nearest camera in viewing direction). We draw a single parameter $t_{\text{target}} \sim \mathcal{U}([t_{\min}, t_{\max}])$ with $t_{\min} < 0$ and $t_{\max} > 1$, shared by rotation and translation:

$$\mathbf{R}_{\text{target}} = \text{Slerp}(\mathbf{R}_m, \mathbf{R}_n, t_{\text{target}}), \quad \mathbf{t}_{\text{target}} = (1 - t_{\text{target}})\,\mathbf{t}_m + t_{\text{target}}\,\mathbf{t}_n, \quad (17)$$

where $\text{Slerp}(\mathbf{R}_m, \mathbf{R}_n, t) = \mathbf{R}_m(\mathbf{R}_m^T\mathbf{R}_n)^t$ interpolates along the geodesic path on $SO(3)$. A value $t_{\text{target}} \in [0, 1]$ yields interpolation (intra-view consistency), while $t_{\text{target}} \notin [0, 1]$ yields extrapolation (beyond the training convex hull); the two regimes differ only in the sampled value.

(3) Parallax Translation. We compute a lateral translation perpendicular to the viewing direction to maximize parallax on near-field objects:

$$\mathbf{t}_{\text{parallax}} = \delta \cdot \mathbf{t}_{\perp,m}, \quad (18)$$

where $\mathbf{t}_{\perp,m}$ is a unit vector perpendicular to viewing direction $\mathbf{R}_m\mathbf{e}_z$ (with $\mathbf{e}_z = [0, 0, 1]^T$) and $\delta \sim \mathcal{N}(0, (0.1r_{\text{scene}})^2)$ with scene radius r_{scene}. For 360° scenes, we instead sample on a spherical shell: $\mathbf{t}_k = \mathbf{c}_{\text{scene}} + r_{\text{mean}}\mathbf{u}$, where $\mathbf{u} \sim \text{Uniform}(\mathbb{S}^2)$.

We **integrate the three components** into a single novel view:

$$\mathbf{R}_k = \mathbf{R}_{\text{target}} \cdot \exp(\theta \cdot \mathbf{n}_{\text{edge},m}), \quad \mathbf{t}_k = \mathbf{t}_{\text{target}} + \mathbf{t}_{\text{parallax}}, \quad (19)$$

where θ controls the occlusion boundary rotation magnitude, and $\exp : \mathbb{R}^3 \to SO(3)$ is the exponential map from axis-angle representation to rotation matrix.

We generate $K = M$ novel poses per iteration—one composite probe $\mathbf{P}_k$ per training camera $\mathbf{P}_m$. We render using both base parameters $\boldsymbol{\Theta}$ and attribute-probed parameters $\tilde{\boldsymbol{\Theta}}$ (from Eq. (14)):

$$\hat{\mathbf{I}}_k(\boldsymbol{\Theta}) = \{\mathbf{C}(\mathbf{p}; \boldsymbol{\Theta}, \mathbf{P}_k)\}_{\mathbf{p}}, \quad \hat{\mathbf{I}}_k(\tilde{\boldsymbol{\Theta}}) = \{\mathbf{C}(\mathbf{p}; \tilde{\boldsymbol{\Theta}}, \mathbf{P}_k)\}_{\mathbf{p}}, \quad (20)$$

where $\mathbf{C}(\cdot)$ is the rasterization function from Eq. (1).

Unlike ASP which supervises against ground truth images, VSP measures rendering consistency between base and probed configurations:

$$\mathcal{L}_{\text{VSP}}(\boldsymbol{\Theta}) = \frac{1}{K}\sum_{k=1}^{K}\left[\ell_1\left(\hat{\mathbf{I}}_k(\boldsymbol{\Theta}), \hat{\mathbf{I}}_k(\tilde{\boldsymbol{\Theta}})\right) + \lambda_{\text{SSIM}}\mathcal{L}_{\text{SSIM}}\left(\hat{\mathbf{I}}_k(\boldsymbol{\Theta}), \hat{\mathbf{I}}_k(\tilde{\boldsymbol{\Theta}})\right)\right]. \quad (21)$$

3.5 Manifold-Guided Optimization

As shown in Figure 3 and Algorithm 1, our complete objective combines reconstruction accuracy, attribute-space fragility, and viewpoint-space consistency:

$$\mathcal{L}_{\text{total}}(\boldsymbol{\Theta}) = \mathcal{L}_{\text{recon}}(\boldsymbol{\Theta}) + \lambda_{\text{ASP}}\mathcal{L}_{\text{ASP}}(\boldsymbol{\Theta}) + \lambda_{\text{VSP}}\mathcal{L}_{\text{VSP}}(\boldsymbol{\Theta}), \quad (22)$$

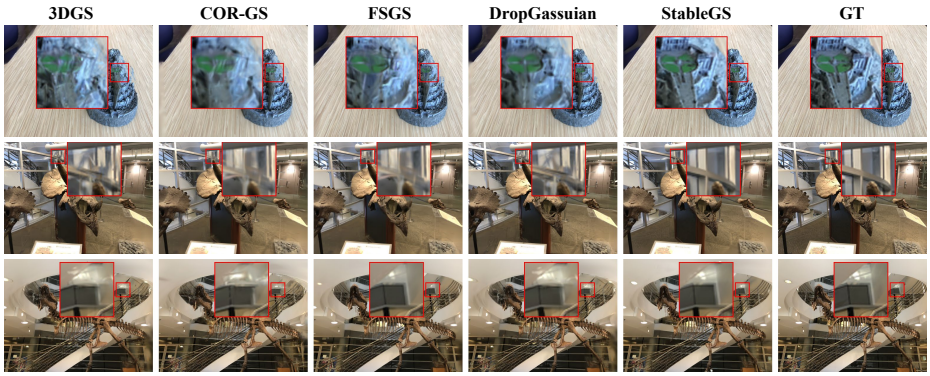

Fig. 4: Qualitative Results on LLFF [30] compared with other methods [16, 36, 57, 61]. Our method achieves better view consistency.

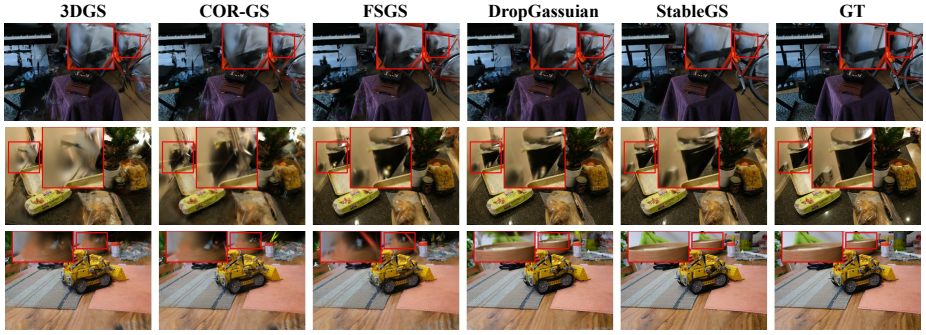

Fig. 5: Qualitative Results on MipNeRF360 [2] compared with other methods [16, 36, 57, 61]. Our method excels in sparse-view scenarios with superior details.

where $\mathcal{L}_{\text{recon}}$ (Eq. (2)) ensures photometric accuracy on training views, $\mathcal{L}_{\text{ASP}}$ (Eq. (15)) measures rendering fragility under attribute probes, and $\mathcal{L}_{\text{VSP}}$ (Eq. (21)) measures viewpoint consistency.

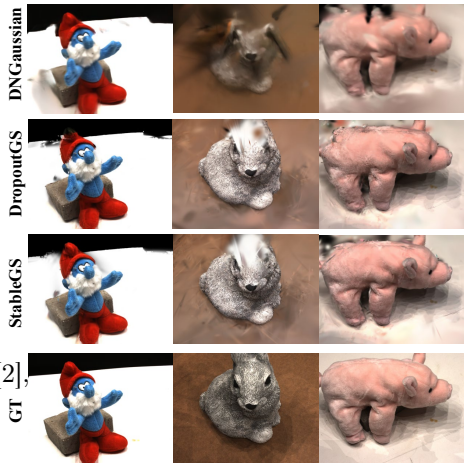

Fig. 6: Qualitative Results on DTU [14].

4 Experiments

Datasets. We evaluate on three benchmarks: LLFF [30], Mip-NeRF360 [2], and DTU [14], following the standard sparse-view protocols of [34, 54, 61]. For LLFF [30] we use 8× downsampling (378 × 504), hold out every 8-th image as a test view, and evenly sample 3/6/9 input views from the remainder. For DTU [14] we use 4× downsampling, the PixelNeRF/RegNeRF split

Table 1: Quantitative comparisons on LLFF [30] with 3, 6, and 9 views. Best, second-best, and third-best results are marked in red, orange, and yellow.

	Methods	3-view			6-view			9-view		
		PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF-based	Mip-NeRF [2]	16.11	0.401	0.460	22.91	0.756	0.213	24.88	0.826	0.170
	DietNeRF [13]	14.94	0.370	0.496	21.75	0.717	0.248	24.28	0.801	0.183
	RegNeRF [34]	19.08	0.587	0.336	23.10	0.760	0.206	24.86	0.820	0.161
	FreeNeRF [54]	19.63	0.612	0.308	23.73	0.779	0.195	25.13	0.827	0.160
	SparseNeRF [48]	19.86	0.624	0.328	-	-	-	-	-	-
3DGS-based	3DGS [16]	16.46	0.440	0.401	21.09	0.699	0.229	23.21	0.785	0.176
	DNGaussian [20]	19.12	0.591	0.294	22.18	0.755	0.198	23.17	0.788	0.180
	DropoutGS [53]	19.23	0.614	0.287	23.35	0.791	0.177	24.33	0.825	0.160
	FSGS [61]	20.43	0.682	0.248	24.09	0.823	0.145	25.31	0.860	0.122
	CoR-GS [57]	20.45	0.712	**0.196**	24.49	0.837	**0.115**	26.06	0.874	0.089
	DropGaussian [36]	20.76	0.713	0.200	24.74	0.837	0.117	26.21	0.874	**0.088**
	StableGS (Ours)	**21.19**	**0.726**	0.208	**25.05**	**0.841**	0.127	**26.47**	**0.885**	0.097

Table 2: Individual impact of each geometric fragility axis across all three datasets. Position probing dominates, recovering ∼84% of the ASP-only gain on LLFF by itself, confirming depth-sorting instability as the primary failure mode. The axes are complementary but target partially overlapping instabilities.

ASP Component	LLFF 3-view			DTU 3-view			Mip-NeRF360 12-view		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Baseline (3DGS)	16.46	0.440	0.401	14.74	0.672	0.249	18.52	0.523	0.415
+ Position only	18.92	0.582	0.312	18.35	0.765	0.189	19.31	0.552	0.391
+ Opacity only	17.85	0.538	0.348	16.98	0.725	0.218	19.02	0.541	0.401
+ Rotation/Scale only	17.52	0.512	0.365	16.42	0.708	0.228	18.89	0.536	0.406
Full ASP+VSP (StableGS)	**21.19**	**0.726**	**0.208**	**20.86**	**0.848**	**0.130**	**19.98**	**0.580**	**0.368**

with input IDs $\{25, 22, 28\}$, and evaluate on the 15 standard scenes within object masks [34]. For Mip-NeRF360 [2] we use 8× downsampling with 12/24 input views.

Implementation. We build on 3DGS [16], training for 10k iterations on LLFF and DTU [14], and 30k on Mip-NeRF360 [2]. The global scaling constant in Eq. (5) is set as $c_\phi = 0.01$; all hyperparameters (λ_{ASP}, λ_{VSP}, c_ϕ) are selected on the **LLFF validation split** and applied *without modification* to DTU and Mip-NeRF360. The Hessian trace estimator (Eq. (6)) uses $B = 10$ samples, updated every 100 iterations. Loss weights: $\lambda_{\text{ASP}} = 1.0$, $\lambda_{\text{VSP}} = 0.5$. We use a single RTX 5090 GPU. We report PSNR, SSIM [49], and LPIPS [58], with DTU [14] metrics computed within object masks.

Baselines. We compare against NeRF methods (Mip-NeRF [1], DietNeRF [13], RegNeRF [34], FreeNeRF [54], SparseNeRF [48]) and 3DGS methods (3DGS [16], DNGaussian [20], DropoutGS [53], FSGS [61], CoR-GS [57], DropGaussian [36]).

4.1 Main Results

LLFF [30]. Table 1 and Figure 4 show results across 3, 6, and 9 views. On 3-view LLFF [30], StableGS achieves 21.19 dB PSNR, outperforming ensemble-based DropGaussian [36] by 0.43 dB (21.19 vs 20.76 dB). Compared to depth-prior methods, we surpass DNGaussian [20] by 2.07 dB without external monocular supervision [40], validating that probe-guided manifold optimization provides stronger geometric constraints than injected priors. Superior LPIPS scores (0.208 vs 0.308 for FreeNeRF [54]) demonstrate that smoother optimization manifold yield more physically plausible geometry.

Methods	12-view			24-view		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS [16]	18.52	0.523	0.415	22.80	0.708	0.276
FSGS [61]	18.80	0.531	0.418	23.70	0.745	0.230
CoR-GS [57]	19.52	0.558	0.418	23.39	0.727	0.271
DropGaussian [36]	19.74	0.577	**0.364**	24.13	**0.762**	**0.225**
StableGS (Ours)	**19.98**	**0.580**	0.368	**24.38**	0.756	0.233

Table 3: Mip-NeRF360 [2] (12 & 24 views). Red / orange / yellow = best/second/third.

DTU [14]. Table 5 and Figure 6 show results on object-centric scenes with 3 views. StableGS achieves state-of-the-art performance (20.86 dB PSNR, 0.848 SSIM, 0.130 LPIPS), outperforming DropoutGS by 0.64 dB and DNGaussian by 1.95 dB without depth supervision. Against MVS-NeRF [3] (18.54 dB), we achieve 2.32 dB improvement through per-scene manifold diagnosis alone.

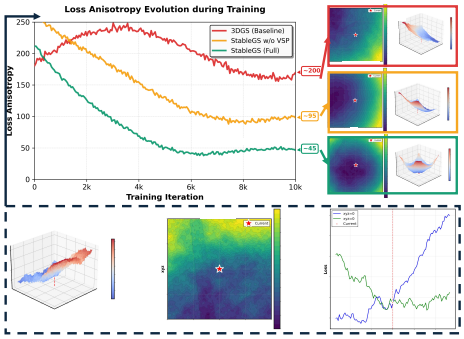

Fig. 7: Optimization manifold diagnosis. Our probes cut the anisotropy index $\mathcal{A}$ (Eq. (3)) by 74%. *Top-right:* final manifolds—StableGS forms a flat basin versus 3DGS's elongated valley. *Bottom:* the initial SfM manifold along the top Hessian eigenvectors (**green** $\mathbf{e}_{\max}$, **blue** $\mathbf{e}_2$), whose sloped geometry motivates ASP/VSP.

Mip-NeRF360 [2]. Table 3 and Figure 5 present results on unbounded 360° scenes. StableGS achieves 19.98 dB (12 views) and 24.38 dB (24 views), surpassing DropGaussian by 0.24 dB and 0.25 dB respectively. Compared to CoR-GS [57] co-regularization, our systematic probing provides more effective geometric constraints (19.98 vs 19.52 dB on 12 views).

4.2 Ablation Studies

Component Analysis. Table 6 presents ablation results on LLFF 3-view. Starting from baseline ($\lambda_{\text{ASP}} = \lambda_{\text{VSP}} = 0$: 16.46 dB), ASP alone ($\lambda_{\text{ASP}} = 1.0$: +2.92 dB) diagnoses attribute fragility on training views, while VSP alone ($\lambda_{\text{VSP}} = 0.5$: +2.29 dB) diagnoses viewpoint fragility on novel poses.

Configuration	Occlusion	Interp/ Rotation	Parallax Extrap	Parallax Trans	LLFF [30] 3-view PSNR↑	SSIM↑	LPIPS↓
Occlusion only	✓				19.87	0.658	0.276
Interpolation only		✓			19.62	0.651	0.287
Parallax only				✓	19.43	0.642	0.295
Occlusion + Interp	✓	✓			20.31	0.682	0.249
Occlusion + Parallax	✓			✓	20.18	0.675	0.258
Interp + Parallax		✓		✓	20.09	0.668	0.263
Full	✓	✓		✓	**21.19**	**0.726**	**0.208**

Table 4: VSP component ablation. Each row tests different combinations of the three geometric probing modes.

Combined StableGS achieves 21.19 dB; on top of ASP, VSP contributes a further +1.81 dB, confirming that effective manifold diagnosis requires probing both attribute and viewpoint spaces. All hyperparameters are selected on the LLFF validation split. Table 6 also confirms they transfer without modification to DTU and Mip-NeRF360, with ±0.1 weight perturbations causing at most 0.13 dB degradation and ±0.01 magnitude perturbations causing at most 0.05 dB degradation.

Geometric Fragility Axes. Table 2 quantifies the individual contribution of each fragility axis identified in Sec. 3.2. Position probing provides the largest single-axis gain (+2.46 dB on LLFF, recovering ∼84% of the +2.92 dB ASP-only gain by itself), confirming depth-sorting instability as the dominant failure in sparse-view 3DGS. Opacity and rotation/scale probing contribute +1.39 dB and +1.06 dB respectively. The axes are complementary but overlapping: the three standalone gains sum to +4.91 dB while combined ASP yields +2.92 dB, indicating they target partially shared instabilities rather than acting independently.

Optimization Manifold Metrics. Figure 7 quantifies manifold properties using the anisotropy index from Eq. (3). Compared to 3DGS (red curve), StableGS (green curve) reduces anisotropy $\mathcal{A}$ by 74%, confirming our hypothesis that geometric probes reshape directionally anisotropic surfaces toward isotropic basins. The top-right insets show the final manifolds: StableGS forms the most pronounced flat basin, contrasting with 3DGS's elongated valley. The bottom row shows the initial SfM-initialized manifold (before training), whose sloped, basin-free geometry motivates ASP/VSP.

Table 5: Quantitative comparison on DTU [14]. Best, second, and third results marked in red, orange, and yellow.

Method	Setting	DTU [14]		
		PSNR ↑	SSIM ↑	LPIPS ↓
SRF [5]	Trained on DTU [14]	15.32	0.671	0.304
PixelNeRF [55]		16.82	0.695	0.270
MVSNeRF [3]		18.63	0.769	0.197
SRF ft [5]	Trained on DTU [14] Fine-tuned per Scene	15.68	0.698	0.281
PixelNeRF ft [55]		18.95	0.710	0.269
MVSNeRF ft [3]		18.54	0.769	0.197
Mip-NeRF [1]	Optimized per Scene	8.68	0.571	0.353
DietNeRF [13]		11.85	0.633	0.314
RegNeRF [34]		18.89	0.745	0.190
FreeNeRF [54]		19.92	0.787	0.182
SparseNeRF [48]		19.55	0.769	0.201
3DGS [16]		14.74	0.672	0.249
DNGaussian [20]		18.91	0.790	0.176
DropoutGS [53]		20.22	0.830	0.150
StableGS (Ours)		**20.86**	**0.848**	**0.130**

Viewpoint Probe Sampling. Table 4 validates our composite probe design. Individual modes yield 19.43–19.87 dB, while pairwise combinations improve results to 20.09–20.31 dB. The full composite probe achieves the best performance (21.19 dB), confirming occlusion boundaries, multi-view interpolation, and depth-parallax consistency provide diagnosis.

4.3 Efficiency Analysis

Figure 8 plots per-iteration PSNR-vs-time trajectories. At the **9.7-min checkpoint**—the same wall-clock budget as vanilla 3DGS—StableGS already reaches **21.16 dB**, exceeding the *converged* result of every baseline (DropGaussian 20.76 dB@11.8 min, CoR-GS 20.45@13.0, FSGS 20.43@12.5, DropoutGS 19.23@11.5, DNGaussian 19.12@11.0); its own final result is 21.19 dB at 13.1 min. Training vanilla 3DGS out to 13.1 min reaches only 16.51 dB, confirming the improvement stems from the reshaped optimization manifold rather than extra compute.

Table 6: Cross-dataset hyperparameter robustness. Blue: decreased value; orange: increased value.

λ_{ASP}	λ_{VSP}	c_ϕ	LLFF		DTU		Mip-NeRF360	
			3-view	6-view	3-view	6-view	12-view	24-view
Component Contributions (PSNR ↑)								
0	0	-	16.46	21.09	14.74	18.32	18.52	22.80
1.0	0	0.01	19.38 (+2.92)	23.58 (+2.49)	18.20 (+3.46)	20.15 (+1.83)	19.12 (+0.60)	23.45 (+0.65)
0	0.5	0.01	18.75 (+2.29)	22.94 (+1.85)	17.85 (+3.11)	19.68 (+1.36)	18.89 (+0.37)	23.21 (+0.41)
1.0	**0.5**	**0.01**	**21.19** (+4.73)	**25.05** (+3.96)	**20.86** (+6.12)	**22.34** (+4.02)	**19.98** (+1.46)	**24.38** (+1.58)
Sensitivity: ±0.1 Loss Weights ($c_\phi = 0.01$, selected on LLFF val. split)								
0.9	0.5	0.01	21.06 (-0.13)	24.92 (-0.13)	20.78 (-0.08)	22.26 (-0.08)	19.89 (-0.09)	24.29 (-0.09)
1.0	**0.5**	**0.01**	**21.19**	**25.05**	**20.86**	**22.34**	**19.98**	**24.38**
1.1	0.5	0.01	21.11 (-0.08)	24.98 (-0.07)	20.82 (-0.04)	22.30 (-0.04)	19.94 (-0.04)	24.33 (-0.05)
1.0	0.4	0.01	21.08 (-0.11)	24.95 (-0.10)	20.81 (-0.05)	22.28 (-0.06)	19.91 (-0.07)	24.31 (-0.07)
1.0	0.6	0.01	21.13 (-0.06)	25.00 (-0.05)	20.84 (-0.02)	22.32 (-0.02)	19.95 (-0.03)	24.35 (-0.03)
Sensitivity: ±0.01 Probe Magnitude ($\lambda_{ASP} = 1.0, \lambda_{VSP} = 0.5$)								
1.0	0.5	0.009	21.14 (-0.05)	25.00 (-0.05)	20.84 (-0.02)	22.31 (-0.03)	19.95 (-0.03)	24.35 (-0.03)
1.0	**0.5**	**0.01**	**21.19**	**25.05**	**20.86**	**22.34**	**19.98**	**24.38**
1.0	0.5	0.011	21.16 (-0.03)	25.02 (-0.03)	20.85 (-0.01)	22.33 (-0.01)	19.96 (-0.02)	24.36 (-0.02)

In terms of cost-benefit, StableGS achieves +4.73 dB over 3DGS at 1.3× training time (4.73/1.3 = 3.64 dB per training-time factor), versus DropGaussian's +4.30 dB at 1.2× (4.30/1.2 = 3.58 dB per factor)—a higher absolute gain and a better efficiency ratio. Critically, inference incurs **zero additional cost**: ASP and VSP are training-only, and the final model is a standard 3DGS representation indistinguishable from the baseline at render time.

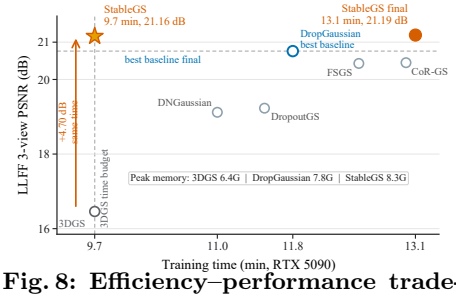

Fig. 8: Efficiency–performance trade-off on LLFF [30] 3-view (single RTX 5090). Per-iteration PSNR-vs-time; the star marks StableGS at the vanilla 3DGS wall-clock budget (9.7 min), already exceeding every baseline's converged result.

5 Conclusion

StableGS reshapes the anisotropic optimization manifold toward isotropy via geometric probing. Identifying three geometric fragility axes, we propose ASP and VSP to target rendering-specific instabilities, yielding +0.33/+0.64/+0.25 dB PSNR on LLFF [30], DTU [14], and Mip-NeRF360 [2] with 74% anisotropy reduction—a principled route to robust sparse-view reconstruction.

Acknowledgements This work was supported by the National Natural Science Foundation of China (62572494, U25B2036, U2368201), the Hunan Provincial Science and Technology Innovation Plan—Key Field Research and Development Program (Grant No. 2024AQ2040), and the High Performance Computing Center of Central South University.

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021). https://doi.org/10.1109/iccv48922.2021.00580
2. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022). https://doi.org/10.1109/cvpr52688.2022.00539
3. Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H.: Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021). https://doi.org/10.1109/iccv48922.2021.01386
4. Cheng, K., Lei, C., Lei, S., Long, X., Lu, C.T., Wang, L., Wang, S., Yin, W.: Dc-gaussian: Improving 3d gaussian splatting for reflective dash cam videos. Advances in Neural Information Processing Systems 37 (2024). https://doi.org/10.52202/079017-3169
5. Chibane, J., Bansal, A., Lazova, V., Pons-Moll, G.: Stereo radiance fields (srf): Learning view synthesis for sparse views of novel scenes. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021). https://doi.org/10.1109/cvpr46437.2021.00782
6. Chng, S.F., Ramasinghe, S., Sherrah, J., Lucey, S.: Gaussian activated neural radiance fields for high fidelity reconstruction and pose estimation. Lecture Notes in Computer Science (2022). https://doi.org/10.1007/978-3-031-19827-4_16
7. Chung, J., Oh, J., Lee, K.M.: Depth-regularized optimization for 3d gaussian splatting in few-shot images. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2024). https://doi.org/10.1109/cvprw63382.2024.00086
8. Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B.: Sharpness-aware minimization for efficiently improving generalization. arXiv (Cornell University) (2020). https://doi.org/10.48550/arxiv.2010.01412
9. Goodfellow, I., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv (Cornell University) (2014). https://doi.org/10.48550/arxiv.1412.6572
10. Gu, J., Liu, L., Wang, P., Theobalt, C.: Stylenerf: A style-based 3d-aware generator for high-resolution image synthesis. arXiv (Cornell University) (2021). https://doi.org/10.48550/arxiv.2110.08985
11. Heo, H., Kim, T., Lee, J., Lee, J., Kim, S.H., Kim, H.J., Kim, J.H.: Robust camera pose refinement for multi-resolution hash encoding. arXiv (Cornell University) (2023). https://doi.org/10.48550/arxiv.2302.01571
12. Hutchinson, M.: A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. Communications in Statistics - Simulation and Computation (1989). https://doi.org/10.1080/03610918908812806
13. Jain, A., Tancik, M., Abbeel, P.: Putting nerf on a diet: Semantically consistent few-shot view synthesis. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021). https://doi.org/10.1109/iccv48922.2021.00583
14. Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanaes, H.: Large scale multi-view stereopsis evaluation. 2014 IEEE Conference on Computer Vision and Pattern Recognition (2014). https://doi.org/10.1109/cvpr.2014.59

15. Jeong, Y., Ahn, S., Choy, C., Anandkumar, A., Cho, M., Park, J.: Self-calibrating neural radiance fields. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021). https://doi.org/10.1109/iccv48922.2021.00579
16. Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics (2023). https://doi.org/10.1145/3592433
17. Lan, L., Shao, T., Lu, Z., Zhang, Y., Jiang, C., Yang, Y.: 3dgs2: Near second-order converging 3d gaussian splatting. In: Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers. pp. 1–10 (2025)
18. Lee, J.C., Rho, D., Sun, X., Ko, J.H., Park, E.: Compact 3d gaussian representation for radiance field. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024). https://doi.org/10.1109/cvpr52733.2024.02052
19. Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T.: Visualizing the loss landscape of neural nets. Repository for Publications and Research Data (ETH Zurich) (2018). https://doi.org/10.3929/ethz-b-000461393
20. Li, J., Zhang, J., Bai, X., Zheng, J., Ning, X., Zhou, J., Gu, L.: Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024). https://doi.org/10.1109/cvpr52733.2024.01963
21. Li, J., Qiu, X., Xu, L., Guo, L., Qu, D., Long, T., Fan, C., Li, M.: Unif2ace: Fine-grained face understanding and generation with unified multimodal models. arXiv preprint arXiv:2503.08120 **3** (2025)
22. Li, M., Xu, X., Fan, H., Zhou, P., Liu, J., Liu, J.W., Li, J., Keppo, J., Shou, M.Z., Yan, S.: Stprivacy: Spatio-temporal privacy-preserving action recognition. In: ICCV (2023)
23. Li, M., Zhou, P., Liu, J.W., Keppo, J., Lin, M., Yan, S., Xu, X.: Instant3d: Instant text-to-3d generation. International Journal of Computer Vision **132**(10), 4456–4472 (2024)
24. Li, Y., Duan, J., Qu, Z.: Tri-robust learning: Robust multi-neural networks against extremely noisy labels (2024). https://doi.org/10.2139/ssrn.4911734
25. Liu, T., Wang, G., Hu, S., Shen, L., Ye, X., Zang, Y., Cao, Z., Li, W., Liu, Z.: Mvsgaussian: Fast generalizable gaussian splatting reconstruction from multi-view stereo. Lecture Notes in Computer Science (2024). https://doi.org/10.1007/978-3-031-72649-1_3
26. Luo, Z.Q., Tseng, P.: Error bounds and convergence analysis of feasible descent methods: a general approach. Annals of Operations Research (1993). https://doi.org/10.1007/bf02096261
27. Martin-Brualla, R., Radwan, N., Sajjadi, M.S.M., Barron, J.T., Dosovitskiy, A., Duckworth, D.: Nerf in the wild: Neural radiance fields for unconstrained photo collections. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021). https://doi.org/10.1109/cvpr46437.2021.00713
28. Mayers, D.F., Golub, G.H., van Loan, C.F.: Matrix computations. Mathematics of Computation (1986). https://doi.org/10.2307/2008107
29. Meng, Q., Chen, A., Luo, H., Wu, M., Su, H., Xu, L., He, X., Yu, J.: Gnerf: Gan-based neural radiance field without posed camera. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021). https://doi.org/10.1109/iccv48922.2021.00629
30. Mildenhall, B., Srinivasan, P.P., Ortiz-Cayon, R., Kalantari, N.K., Ramamoorthi, R., Ng, R., Kar, A.: Local light field fusion: Practical view synthesis with prescrip-

tive sampling guidelines. arXiv (Cornell University) (2019). https://doi.org/10.48550/arxiv.1905.00889
31. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Lecture Notes in Computer Science (2020). https://doi.org/10.1007/978-3-030-58452-8_24
32. Mądry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. arXiv (Cornell University) (2017). https://doi.org/10.48550/arxiv.1706.06083
33. Neff, T., Stadlbauer, P., Parger, M., Kurz, A., Mueller, J.H., Chaitanya, C.R.A., Kaplanyan, A., Steinberger, M.: Donerf: Towards real-time rendering of compact neural radiance fields using depth oracle networks. Computer Graphics Forum (2021). https://doi.org/10.1111/cgf.14340
34. Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S.M., Geiger, A., Radwan, N.: Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022). https://doi.org/10.1109/cvpr52688.2022.00540
35. Paliwal, A., Ye, W., Xiong, J., Kotovenko, D., Ranjan, R., Chandra, V., Kalantari, N.K.: Coherentgs: Sparse novel view synthesis with coherent 3d gaussians. Lecture Notes in Computer Science (2024). https://doi.org/10.1007/978-3-031-73404-5_2
36. Park, H., Ryu, G., Kim, W.: Dropgaussian: Structural regularization for sparse-view gaussian splatting. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025). https://doi.org/10.1109/cvpr52734.2025.02012
37. Pehlivan, H., Camiletto, A.B., Foo, L.G., Habermann, M., Theobalt, C.: Matrix-free second-order optimization of gaussian splats with residual sampling. In: 2026 International Conference on 3D Vision (3DV). pp. 18–28. IEEE (2026)
38. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision (2021). https://doi.org/10.48550/arxiv.2103.00020
39. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021). https://doi.org/10.1109/iccv48922.2021.01196
40. Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., Koltun, V.: Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022). https://doi.org/10.1109/tpami.2020.3019967
41. Roessle, B., Barron, J.T., Mildenhall, B., Srinivasan, P.P., Niebner, M.: Dense depth priors for neural radiance fields from sparse input views. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022). https://doi.org/10.1109/cvpr52688.2022.01255
42. Sajjadi, M.S., Meyer, H., Pot, E., Bergmann, U., Greff, K., Radwan, N., Vora, S., Lucic, M., Duckworth, D., Dosovitskiy, A., Uszkoreit, J., Funkhouser, T., Tagliasacchi, A.: Scene representation transformer: Geometry-free novel view synthesis through set-latent scene representations. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022). https://doi.org/10.1109/cvpr52688.2022.00613

43. Shen, J., Agudo, A., Moreno-Noguer, F., Ruiz, A.: Conditional-flow nerf: Accurate 3d modelling with reliable uncertainty quantification. Lecture Notes in Computer Science (2022). https://doi.org/10.1007/978-3-031-20062-5_31
44. Sun, C., Triantafyllou, T., Makris, A., Drmač, M., Xu, K., Mai, L., Marina, M.K.: Ph-dropout: Practical epistemic uncertainty quantification for view synthesis. arXiv (Cornell University) (2024). https://doi.org/10.48550/arxiv.2410.05468
45. Tan, S., Qiu, X., Shu, Y., Xu, G., Xu, L., Xu, X., Zhuang, H., Li, M., Yu, F.: WMarkGPT: Watermarked image understanding via multimodal large language models. In: Proceedings of the 42nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 267, pp. 58621–58636. PMLR (13–19 Jul 2025)
46. Tonderski, A., Lindström, C., Hess, G., Ljungbergh, W., Svensson, L., Petersson, C.: Neurad: Neural rendering for autonomous driving. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024). https://doi.org/10.1109/cvpr52733.2024.01411
47. Wan, L., Zeiler, M.D., Zhang, S., Lecun, Y., Fergus, R.: Regularization of neural networks using dropconnect. International review of cytology (2013). https://doi.org/10.1016/s0074-7696(08)60205-3
48. Wang, G., Chen, Z., Loy, C.C., Liu, Z.: Sparsenerf: Distilling depth ranking for few-shot novel view synthesis. 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (2023). https://doi.org/10.1109/iccv51070.2023.00832
49. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
50. Wei, H., Feng, L., Chen, X., An, B.: Combating noisy labels by agreement: A joint training method with co-regularization. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020). https://doi.org/10.1109/cvpr42600.2020.01374
51. Wu, Z., Li, X., Peng, J., Lu, H., Cao, Z., Zhong, W.: Dof-nerf: Depth-of-field meets neural radiance fields. Proceedings of the 30th ACM International Conference on Multimedia (2022). https://doi.org/10.1145/3503161.3548088
52. Wynn, J., Turmukhambetov, D.: Diffusionerf: Regularizing neural radiance fields with denoising diffusion models. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023). https://doi.org/10.1109/cvpr52729.2023.00407
53. Xu, Y., Wang, L., Chen, M., Ao, S., Li, L., Guo, Y.: Dropoutgs: Dropping out gaussians for better sparse-view rendering. 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025). https://doi.org/10.1109/cvpr52734.2025.00074
54. Yang, J., Pavone, M., Wang, Y.: Freenerf: Improving few-shot neural rendering with free frequency regularization. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023). https://doi.org/10.1109/cvpr52729.2023.00798
55. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021). https://doi.org/10.1109/cvpr46437.2021.00455
56. Yuan, R., Gower, R.M., Lazaric, A.: A general sample complexity analysis of vanilla policy gradient. arXiv (Cornell University) (2021). https://doi.org/10.48550/arxiv.2107.11433

57. Zhang, J., Li, J., Yu, X., Huang, L., Gu, L., Zheng, J., Bai, X.: Cor-gs: Sparse-view 3d gaussian splatting via co-regularization. Lecture Notes in Computer Science (2024). https://doi.org/10.1007/978-3-031-73232-4_19
58. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018). https://doi.org/10.1109/cvpr.2018.00068
59. Zhao, B., Dehmamy, N., Walters, R., Yu, R.: Symmetry teleportation for accelerated optimization. arXiv (Cornell University) (2022). https://doi.org/10.48550/arxiv.2205.10637
60. Zhao, C., Wang, X., Zhang, T., Javed, S., Salzmann, M.: Self-ensembling gaussian splatting for few-shot novel view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4940–4950 (2025)
61. Zhu, Z., Fan, Z., Jiang, Y., Wang, Z.: Fsgs: Real-time few-shot view synthesis using gaussian splatting. Lecture Notes in Computer Science (2024). https://doi.org/10.1007/978-3-031-72933-1_9
62. Zwicker, M., Pfister, H., van Baar, J., Gross, M.: Ewa volume splatting. Proceedings Visualization, 2001. VIS '01. (nd). https://doi.org/10.1109/visual.2001.964490