SynGS: Synergistic Continual Learning and Change Detection with Gaussian Splatting

Rui Huang[1] **Haojie Tao**[1] **Sen Gao**[1] **Qing Guo**[2*]

[1] Civil Aviation University of China, Tianjin, China [2] Nankai University, Tianjin, China
* Corresponding author · tsingqguo@ieee.org

Motivation

Revisited scenes may contain inserted, removed, or rearranged objects. Shadows, reflections, and exposure shifts can mimic these structural changes in the images.

Fine-tuning 3D Gaussian Splatting (3DGS) [1] can erase earlier content. Fixed static–dynamic labels lock in early errors. SynGS updates geometry and change estimates together so each can improve the other.

Method

From a pre-trained 3DGS scene and new posed views, SynGS produces an updated scene, multi-view change masks, and a compact record for replay.

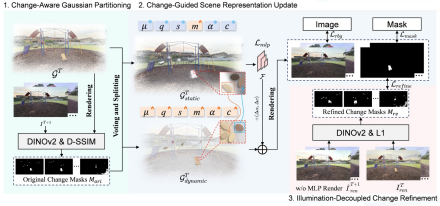


Figure 2. SynGS initializes change cues in 3D, updates scene parameters and change logits jointly, and then refines the masks using illumination-decoupled renders.

Multi-view Initialization

Compare prior renders with new views. Intersect normalized DINOv2 [2] feature differences and 1 − SSIM, both above 0.5. A Gaussian starts as dynamic if more than half its visible views vote for change.

Logit-driven State Transitions

Reassign Gaussians with outlier logits, capped at 1% of each group per transition step. Densification and pruning modify only the dynamic set.

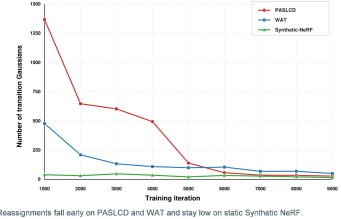


Figure S6. Reassignments fall early on PASLCD and WAT and stay low on static Synthetic NeRF.

Joint Gaussian Optimization

Each Gaussian learns a change probability from a logit. High-probability Gaussians explain changes through standard 3DGS updates. The geometry of Gaussians assigned to the static set stays fixed while a lightweight MLP predicts bounded RGB and opacity residuals.

$$m_i = \sigma(\ell_i), \quad \widetilde{c}_i(d) = c_i(d) + (1 - m_i)\,\Delta c_i, \quad \widetilde{\alpha}_i = \alpha_i + (1 - m_i)\,\Delta\alpha_i$$

The gate scales RGB and opacity corrections (Δc and Δα) by static confidence. Joint training combines RGB reconstruction, mask supervision, residual regularization, and probability binarization. The initial masks upweight the RGB L1 loss in likely changed pixels.

Mask Rendering and Refinement

$$\widehat{M}(p) = \sum_{i=1}^{K} T_i(p)\,\alpha_i(p)\,m_i$$

At pixel p, front-to-back transmittance T and uncorrected opacity α weight each visible Gaussian's change probability.

After the joint update, render with the illumination MLP disabled. Compare this render with the previous scene using DINOv2 features and pixelwise L1 differences, then refine only the change logits with geometry and appearance frozen.

Scene-history Replay

Reuse a shared static backbone and store timestep-specific dynamic Gaussians, static additions, and illumination weights. The saved logits also allow historical masks to be rendered again.

Experiments

Experimental Setup

PASLCD [3] contains 10 indoor and outdoor pairs with real changes and unconstrained camera paths. WAT [4] tests temporal appearance variation. Each Synthetic NeRF scene is split into 10 sequential subsets. SynGS uses 15k iterations per timestep.

Multi-view Change Localization

Method	F1 ↑	mIoU ↑
Initial masks	0.386	0.264
GeSCD [7]	0.611	0.477
MV3DCD [3]	0.612	0.461
SynGS	**0.649**	**0.500**

Table 1. PASLCD. Selected baselines include the strongest competing method on each metric.

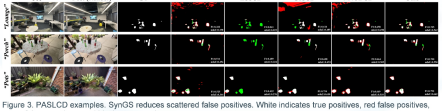


Figure 3. PASLCD examples. SynGS reduces scattered false positives. White indicates true positives, red false positives, green false negatives, and black true negatives.

Continual Scene Updating

SynGS leads the listed baselines in PSNR and rendering speed on both real-world datasets. On static Synthetic NeRF it reaches 33.07 dB at 934 FPS. CLNeRF attains 33.73 dB at 1 FPS.

Method	PASLCD PSNR ↑	PASLCD FPS ↑	WAT PSNR ↑	WAT FPS ↑
3DGS [1]	22.87	118	18.29	55
CLNeRF [4]	26.00	<1	20.94	<1
4DGS [5]	27.64	66	23.38	9
CL-Splats [6]	23.84	391	19.09	195
SynGS	**29.69**	**744**	**26.93**	**565**

Table 3. PSNR in dB. PASLCD is evaluated at 4× downsampling and WAT at native resolution.

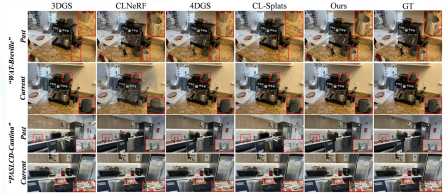


Figure 4. Selected past and current views from WAT–Breville and PASLCD–Cantina. Insets compare the recovery of changed objects and the retention of detail during replay.

WAT replay retains 27.70 dB on T1 after updating through T4, compared with 28.43 dB immediately after T1 training.

Change Detection Ablation

Configuration	F1 ↑	mIoU ↑
Initial masks	0.386	0.264
Joint update without refinement	0.630	0.470
SynGS with refinement	**0.649**	**0.500**

Table 2. On PASLCD, joint optimization accounts for most of the gain, with a further improvement after illumination-decoupled refinement.

Conclusion

Jointly learning change probabilities improves reconstruction and localization while retaining fast rendering and enabling replay.

Complex local lighting, tiny or occluded changes, and severely incomplete or misaligned initial masks remain difficult.

References

[1] Kerbl et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM TOG, 2023.
[2] Oquab et al. DINOv2: Learning Robust Visual Features without Supervision. 2023.
[3] Galappaththige et al. Multi-View Pose-Agnostic Change Localization with Zero Labels. CVPR, 2025.
[4] Cai and Müller. CLNeRF: Continual Learning Meets NeRF. ICCV, 2023.
[5] Wu et al. 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering. CVPR, 2024.
[6] Ackermann et al. CL-Splats: Continual Learning of Gaussian Splatting with Local Optimization. ICCV, 2025.
[7] Kim and Kim. Towards Generalizable Scene Change Detection. CVPR, 2025.