SAGE-GS: SPATIALLY WEIGHTED EVENT SUPERVISION AND SPARSE TEMPORAL GATING FOR DYNAMIC GAUSSIAN SPLATTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Event-guided dynamic reconstruction combines RGB appearance with high-temporal-resolution event observations to recover scene dynamics between sparse image frames. In deformable Gaussian reconstruction, events provide inter-frame supervision, while a deformation network models time-varying geometry. However, events can arise from object motion, noise, illumination changes, or camera motion across static textures, so response magnitude alone does not establish their reliability or relevance to scene dynamics. Geometric deformation also leaves the base opacity of each Gaussian unchanged, limiting direct control over when that primitive contributes to rendering. We propose SAGE-GS to address these two issues through spatial event-supervision allocation and temporal opacity modulation. The supervision branch fuses RGB and event features to predict weights for event residuals, together with a dynamic-region prior, soft event-density weights, and a separate amplitude-based reweighting factor. These factors use appearance and event-response information to adjust the relative supervision assigned to different image regions. The temporal branch gives each dynamic Gaussian a learnable temporal center, a fixed-scale temporal kernel, and an individual gating strength that interpolates between unsuppressed base opacity and temporal-kernel modulation. Zero initialization, projection onto the unit interval, and sparsity regularization encourage selective temporal suppression while allowing some Gaussians to retain their base opacity throughout the sequence. Together, the branches regulate where event observations influence optimization and when dynamic Gaussian primitives contribute to rendering. Experiments on eight synthetic and four real dynamic scenes show average PSNR gains of approximately $0.52\,\mathrm{dB}$ and $0.24\,\mathrm{dB}$, respectively, over E-D3DGS.

1 INTRODUCTION

Reconstructing a dynamic scene requires estimating appearance and time-varying geometry from incomplete observations (Yang et al., 2024; Wu et al., 2024). Event cameras measure brightness changes at high temporal resolution and can constrain motion between RGB frames (Gallego et al., 2022; Bhattacharya et al., 2024). E-D3DGS (Xu et al., 2025) uses these measurements to supervise deformable Gaussians, providing a basis for joint RGB–event reconstruction.

Event residuals differ in their relevance to scene motion. Object motion, camera motion across static textures, and illumination changes can all produce events (Gallego et al., 2022); a large response is therefore not, by itself, evidence of a useful dynamic constraint. A weighting rule based only on event magnitude cannot distinguish these sources. RGB appearance supplies context about the observed structures, while a rendered dynamic-region prior indicates where the current model places moving content. We investigate whether these complementary cues can help allocate event supervision without discarding weak event responses. A separate modeling choice concerns how long each Gaussian should contribute. Deformation changes its geometry, but time-dependent opacity also allows its contribution to diminish away from a temporal center (Li et al., 2024; Wang et al., 2025). As objects rotate or deform, a primitive that fits local structure at one time may be less useful at another. Adjusting its opacity provides an additional degree of freedom alongside geometric motion.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

The choice is then how much temporal suppression to apply to each primitive, while retaining those that remain useful throughout the sequence.

SAGE-GS couples spatial event weighting with sparse temporal opacity gating. Inspired by the RGB–event representation in CompEvent (Zhong et al., 2026), the supervision branch predicts residual weights from fused observation features and detached reconstruction errors. Local event responses, a dynamic-region prior, and amplitude reweighting further shape these weights. The temporal branch interpolates between constant base opacity and a Gaussian temporal kernel for each dynamic primitive. Gate strengths start at zero, are projected to $[0, 1]$, and receive a sparsity penalty. This formulation lets the optimizer retain unsuppressed primitives alongside temporally localized ones.

Across eight synthetic and four real scenes, the reported mean PSNR gains over E-D3DGS are 0.52 and $0.24\,\mathrm{dB}$. Supervision weighting accounts for most of the synthetic gain; gating adds $0.08\,\mathrm{dB}$ on average, with regressions in some scenes. Our contributions are:

1. A spatial event-supervision scheme that combines learned RGB–event residual weights with local event-response and dynamic-region factors. This formulation uses appearance and event-response cues to redistribute supervision across image regions while retaining constraints in weak-response areas.

2. An integration of per-Gaussian temporal opacity gating with deformable RGB–event reconstruction, using zero-initialized projected gate strengths and sparse regularization. This design allows dynamic Gaussians to retain persistent contributions or undergo selective temporal suppression alongside geometric deformation.

2 Related Work

2.1 Dynamic Gaussian Representations and Temporal Contribution Modeling

Dynamic Gaussian representations describe scene motion by modeling Gaussian primitive attributes as functions of time. Deformable-3DGS (Yang et al., 2024) and 4D-GS (Wu et al., 2024) learn time-dependent deformations of Gaussian primitives and represent geometric changes by adjusting their positions, rotations, and scales.

Time-dependent opacity controls when a Gaussian contributes to an image. Spacetime Gaussian Splatting (Li et al., 2024) and FreeTimeGS (Wang et al., 2025) combine temporal opacity with motion modeling. FreeTimeGS++ (Lee et al., 2026) uses a persistence gate to mix persistent and temporally localized contributions. We build on this class of gating functions rather than claim a new temporal kernel. Our focus is their integration with event-supervised deformation and the optimization of sparse suppression strengths.

2.2 Event-Assisted Scene Reconstruction and Cross-Modal Fusion

Event observations provide complementary constraints for scene reconstruction by measuring brightness changes between frames. Early methods mainly incorporated event measurements into neural radiance fields. Ev-NeRF, E-NeRF, and EventNeRF learn scene radiance representations from event observations, while E2NeRF further combines blurry RGB images with event measurements. EvDNeRF extends event-based radiance fields to dynamic scenes (Hwang et al., 2023; Klenk et al., 2023; Rudnev et al., 2023; Qi et al., 2023; Bhattacharya et al., 2024). With the development of 3D Gaussian Splatting, Event-3DGS, E-3DGS, and Event3DGS introduce event-consistency constraints into explicit Gaussian representations (Han et al., 2024; Yin et al., 2025; Xiong et al., 2025).

For dynamic scenes, E-D3DGS (Xu et al., 2025) combines static–dynamic Gaussian decomposition, deformation modeling, and event supervision, providing a direct foundation for RGB–event deformable Gaussian reconstruction. ERF-GS (Bai et al., 2026) considers scenarios with disjoint RGB and event viewpoints. It introduces an event-assisted regularized loss and an event-guided Gaussian densification strategy into the optimization and densification stages, respectively, while reducing its reliance on paired RGB supervision.

Our study concerns spatial event-loss weighting within the E-D3DGS reconstruction framework. We combine learned weights with observation-derived factors and examine their interaction with per-Gaussian temporal opacity.

CompEvent (Zhong et al., 2026) fuses RGB and event features through complex-valued operations for low-light video enhancement and deblurring. We adopt its real/imaginary feature interpretation to construct magnitude and normalized-component features for residual weighting.

3 METHOD

3.1 OVERALL FRAMEWORK AND BASE REPRESENTATION

Given timestamped RGB images, camera parameters, and corresponding events, SAGE-GS learns a deformable Gaussian scene representation for view and time synthesis. We retain the static–dynamic decomposition and event residual of E-D3DGS. Figure 1 summarizes the added supervision-weighting and temporal-gating branches. At a sampled time, the deformation network first determines dynamic Gaussian geometry, and temporal gates modulate the corresponding opacities before rendering. The rendered brightness change defines an event residual. The supervision branch then assigns spatial weights to that residual using observation features and the current reconstruction state. Both branches are trained through the reconstruction objective.

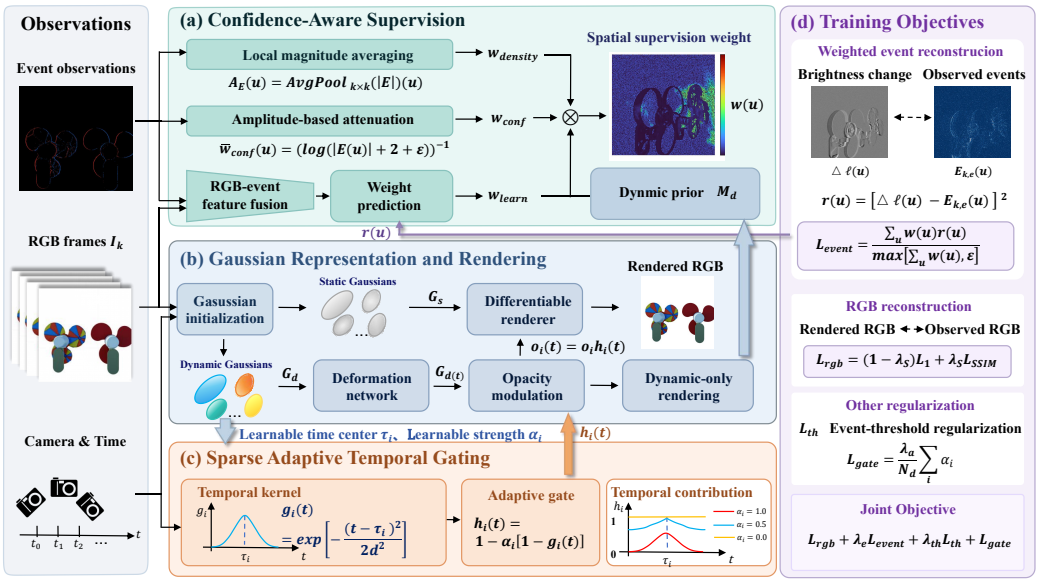


Figure 1: Overall framework of SAGE-GS.

Base Representation. Let $\mathcal{G}_s$ and $\mathcal{G}_d$ denote the static and dynamic Gaussian sets, respectively. Each Gaussian is represented by its position, rotation, scale, appearance parameters, and base opacity o_i. The deformation network adjusts the position, rotation, and scale of dynamic Gaussians according to their spatial locations and time, while their base opacities remain time-invariant. On top of this geometric deformation, we introduce a temporal opacity gate for dynamic Gaussians. Static Gaussians are not subject to deformation or temporal gating.

Event Residual. Given an observed RGB frame I_k at time t_k, we render an image $\hat{I}(t_e)$ at a sampled event time t_e. Let $\Delta\ell$ denote the brightness change between the two images computed according to their temporal order, and let $E_{k,e}$ denote the accumulated brightness change obtained from event counts and learnable event thresholds over the same time interval. The two quantities follow the same temporal direction. The event residual at pixel $\mathbf{u}$ is defined as

$$r(\mathbf{u}) = \left[\Delta\ell(\mathbf{u}) - E_{k,e}(\mathbf{u})\right]^2. \tag{1}$$

Based on this residual, we introduce spatial weights to adjust the supervision strength assigned to different pixels.

3.2 RGB–Event Fusion for Dynamic-Aware Supervision Weighting

Cross-Modal Features and Weight Prediction. Event supervision should consider both the event response and the RGB appearance of the corresponding region. Inspired by the complex-valued cross-modal representation in CompEvent, we treat the RGB feature F_R and event feature F_E extracted by convolutional encoders as the real and imaginary parts of $Z = F_R + \mathrm{i}F_E$, respectively. We then compute their joint magnitude and normalized components:

$$A = \sqrt{F_R^2 + F_E^2 + \epsilon}, \qquad S = \frac{F_E}{\max(A, \epsilon)}, \qquad C = \frac{F_R}{\max(A, \epsilon)}. \tag{2}$$

The magnitude A measures the joint feature response; S and C encode each modality relative to that magnitude. A single-step GRU maps $[A, S, C]$ to a fused feature H at each pixel, with its hidden state initialized for each call. Thus, it acts as a gated feature mapper without carrying state across frames. Separating magnitude from normalized components lets the mapper use both overall feature strength and the relative response of the two modalities. The prediction head ψ receives H and the detached residual r, so its inputs describe the observation as well as the current local fitting error. Detaching r blocks the indirect gradient path from residual-dependent weight prediction to the reconstruction:

$$w_{\mathrm{learn}} = w_{\min} + (1 - w_{\min})\sigma\big(\psi\big([H, \mathrm{sg}(r)]\big)\big)\,. \tag{3}$$

Here, $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation, and $w_{\min} > 0$ is the lower bound of the predicted weight.

Auxiliary Weights and Event Loss. We multiply learned weights by three auxiliary factors. The soft event-response factor w_{density} uses a normalized local average of event magnitude. The dynamic prior M_d uses a smoothed render of the dynamic Gaussian set with unit colors and a black background, with gradients detached. Both have positive lower bounds, so weak-response regions retain supervision. This opacity-based prior is not a ground-truth segmentation label. Finally, w_{conf} decreases with event magnitude. Despite the subscript, it is an amplitude-reweighting heuristic, not a calibrated estimate of reliability. The resulting weight and normalized event loss are

$$w(\mathbf{u}) = w_{\mathrm{learn}}(\mathbf{u})\, w_{\mathrm{density}}(\mathbf{u})\, M_d(\mathbf{u})\, w_{\mathrm{conf}}(\mathbf{u}), \tag{4}$$

$$\mathcal{L}_{\mathrm{event}} = \frac{\sum_{\mathbf{u}} w(\mathbf{u}) r(\mathbf{u})}{\max\left(\sum_{\mathbf{u}} w(\mathbf{u}), \epsilon\right)}. \tag{5}$$

Normalization makes the loss depend on relative spatial weights. When the denominator exceeds the numerical floor, scaling every weight by the same positive constant leaves the loss unchanged. The branch therefore changes which pixels influence fitting, rather than merely rescaling the event-loss coefficient. The local event-response factor favors regions with event activity, while the dynamic prior supplies information from the current scene decomposition. Their soft floors retain constraints outside these regions, where the prior may be incomplete or event activity may be low. The positive floors keep every pixel in the weighted reconstruction objective. For a fixed set of weights, the coefficient on each squared residual is its weight divided by the total weight. Increasing one region's relative weight therefore increases its share of the event objective.

Weight construction. We use 5×5 average pooling $\mathcal{P}_5$ and normalize each map by its spatial maximum: $\mathcal{N}(X) = X / \max(\max_{\mathbf{u}} X(\mathbf{u}), \epsilon)$. For the unit-color dynamic render J_d and its luminance $\ell(J_d)$, the auxiliary factors are

$$\begin{aligned} w_{\mathrm{density}} &= a + (1 - a)\mathcal{N}(\mathcal{P}_5(|E_{k,e}|)), \\ M_d &= b + (1 - b)\mathcal{N}(\mathcal{P}_5(\ell(J_d))), \\ w_{\mathrm{conf}} &= \mathrm{clip}\left(\frac{q}{\max(\mathrm{mean}_{\mathbf{u}}\, q, \epsilon)}, c_{\min}, c_{\max}\right), \quad q = \frac{1}{\log(|E_{k,e}| + 2 + \epsilon)}. \end{aligned} \tag{6}$$

Here, a and b set the soft floors, and $c_{\min}, c_{\max}$ bound amplitude reweighting. These factors are detached from the gradient graph. The event-response and amplitude factors act in opposite directions on event magnitude; their product therefore need not increase monotonically with that magnitude.

3.3 Per-Gaussian Adaptive Temporal Gating and Sparse Optimization

Temporal Gating. For each dynamic Gaussian, we learn a temporal center τ_i and a suppression strength $\alpha_i \in [0, 1]$. A fixed-width kernel modulates its base opacity:

$$g_i(t) = \exp\left[-\frac{(t-\tau_i)^2}{2d^2}\right], \tag{7}$$

$$h_i(t) = 1 - \alpha_i\big(1 - g_i(t)\big), \qquad o_i(t) = o_i h_i(t). \tag{8}$$

The fixed scale d controls the decay rate of the kernel as time moves away from its center. When $\alpha_i = 0$, the Gaussian retains its base opacity, while its geometric motion is still described by the deformation network. When $\alpha_i = 1$, its opacity is fully modulated by the temporal kernel; intermediate values produce partial temporal suppression. At $t = \tau_i$, the kernel equals one and the base opacity is preserved for every gate strength. Far from the center, the contribution approaches $1 - \alpha_i$. Thus, a nonzero gate does not necessarily make a primitive transient: small strengths retain most of its contribution even outside the kernel peak.

Sparse Optimization. We initialize $\alpha_i = 0$ and project it to $[0, 1]$ after each optimizer update. This preserves the ungated representation at initialization and permits exact zeros during training. To discourage suppression across all primitives, we penalize the mean gate strength:

$$\mathcal{L}_{\text{gate}} = \lambda_\alpha \frac{1}{N_d} \sum_{i=1}^{N_d} \alpha_i. \tag{9}$$

Here, N_d is the number of dynamic Gaussians and λ_α controls the penalty. The reconstruction loss competes with this penalty when a nonzero gate reduces fitting error. Some dynamic Gaussians can already follow local structure through deformation, so imposing full kernel modulation on every primitive would unnecessarily restrict their contributions. Starting from zero instead makes the initial render independent of the temporal centers. As suppression becomes active, the centers determine where each affected primitive retains its full opacity. Gate sparsity describes opacity modulation; it does not imply fewer Gaussian primitives or faster rendering. The opacity derivative with respect to the temporal center is proportional to the gate strength. At zero strength, this path contributes no gradient to the center; once the strength becomes positive, the center can be adjusted through reconstruction. The strength and center thus control different aspects of the same temporal modulation.

3.4 Joint Optimization

We jointly optimize the RGB reconstruction loss, the weighted event loss, the event-threshold regularization inherited from E-D3DGS, and the gating sparsity term:

$$\mathcal{L} = \mathcal{L}_{\text{RGB}} + \lambda_e \mathcal{L}_{\text{event}} + \lambda_{\text{th}} \mathcal{L}_{\text{th}} + \mathcal{L}_{\text{gate}}. \tag{10}$$

Here, $\mathcal{L}_{\text{RGB}}$ combines the pixel-wise L_1 loss and SSIM, while $\mathcal{L}_{\text{th}}$ regularizes the learnable event thresholds. The coefficients λ_e and λ_{th} control the strengths of the corresponding loss terms. Spatial weighting acts on the event residual, whereas temporal opacity influences every image rendered from the dynamic Gaussians. Consequently, RGB supervision also constrains the gates, including at observed frame times. The gate penalty balances this fitting objective against the amount of temporal suppression.

4 Experiments

Dataset. We evaluate our method on the eight synthetic dynamic scenes and four real-world dynamic scenes provided by E-D3DGS (Xu et al., 2025), following its training and testing splits.

Baselines. We compare SAGE-GS with two event-assisted dynamic Gaussian reconstruction methods: E-D3DGS and ERF-GS. E-D3DGS serves as the direct baseline of our method. ERF-GS incorporates event constraints into both the optimization and Gaussian densification stages and is

designed for scenarios with misaligned RGB and event viewpoints. Comparisons with ERF-GS are restricted to synthetic scenes.

Real-world settings. We rerun E-D3DGS to obtain its results on the real-world scenes. In the public E-D3DGS code version used in our experiments, we did not find real-scene configurations for `dynamic_radius_scale` and `buffer_radius_scale`. We therefore set these parameters for the real-world scenes and use identical values for E-D3DGS and SAGE-GS to ensure consistent dynamic/static decomposition settings.

4.1 Quantitative Comparison of Reconstruction Quality

Table 1: Full-image reconstruction on eight synthetic and four real scenes. Higher PSNR and SSIM and lower LPIPS are better. S and T denote supervision weighting alone and temporal gating alone. Bold marks the best displayed value in each column.

Method	fan			lego			celestial_globe			hotdog		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
E-D3DGS	30.18	0.964	0.025	31.85	0.967	0.018	28.83	0.976	0.020	36.15	0.974	0.013
ERF-GS	26.00	0.940	0.025	31.14	0.964	0.022	24.68	0.960	0.025	**38.94**	**0.988**	**0.007**
Ablation (S)	30.21	0.963	0.025	**32.29**	**0.969**	**0.015**	29.78	0.983	0.014	36.27	0.974	0.013
Ablation (T)	30.36	0.963	0.025	31.97	0.968	0.016	28.97	**0.997**	0.018	36.19	0.974	0.013
SAGE-GS	**30.58**	**0.966**	**0.023**	32.02	0.968	0.016	**30.29**	0.984	**0.013**	36.14	0.974	0.013
Method	man			materials			waterwheel			music_box		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
E-D3DGS	28.21	0.943	0.037	38.02	0.993	0.003	28.47	0.950	0.033	30.78	0.963	0.039
ERF-GS	25.12	0.920	0.026	39.00	**0.995**	**0.002**	27.31	0.947	**0.019**	27.17	0.956	0.025
Ablation (S)	**29.74**	**0.962**	**0.022**	38.80	0.994	**0.002**	28.18	**0.952**	0.031	30.79	0.965	0.028
Ablation (T)	27.27	0.937	0.046	37.96	0.993	0.003	**28.55**	0.951	0.031	30.85	0.964	0.029
SAGE-GS	29.03	0.956	0.023	**39.03**	**0.995**	**0.002**	28.53	0.950	0.030	**31.06**	**0.967**	**0.024**
Method	Excavator			Jeep			Flowers			Eagle		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
E-D3DGS	29.38	0.904	0.128	27.78	0.850	0.125	**26.733**	0.875	0.113	29.23	**0.886**	0.117
Ablation (S)	**29.72**	0.905	0.105	27.00	**0.853**	0.112	26.57	**0.877**	**0.109**	**30.00**	0.881	0.103
Ablation (T)	28.52	0.905	0.133	27.71	0.849	0.127	26.58	0.875	0.119	**30.00**	0.881	0.103
SAGE-GS	**29.72**	**0.907**	**0.103**	**27.92**	0.851	**0.109**	26.45	0.874	0.116	**30.00**	0.885	**0.096**

Table 1 reports full-image reconstruction results, with averages weighted equally across scenes. On the eight synthetic scenes, SAGE-GS achieves the best mean PSNR, SSIM, and LPIPS among the three compared methods, improving mean PSNR over E-D3DGS by $0.52\,\mathrm{dB}$. Gains are larger on *celestial_globe* and *materials* but marginal on *waterwheel* and *hotdog*, indicating that the benefits of spatial event weighting and temporal opacity modulation are scene-dependent.

SAGE-GS exceeds ERF-GS in PSNR on seven synthetic scenes, although ERF-GS performs better on all three metrics on *hotdog*. This difference may reflect their use of event information: SAGE-GS emphasizes supervision allocation and temporal contribution control, whereas ERF-GS also adjusts primitive distribution through event-guided densification. These mechanisms may therefore suit different reconstruction demands.

On the four real scenes, SAGE-GS improves mean PSNR over E-D3DGS by $0.24\,\mathrm{dB}$ and reduces mean LPIPS by 0.0148, with little change in SSIM. The improvement is therefore more evident in pixel fidelity and perceptual similarity. On *Eagle*, the $0.77\,\mathrm{dB}$ PSNR gain and 0.021 LPIPS reduction accompany a slight SSIM decrease, illustrating that reconstruction gains are not uniform across evaluation metrics.

4.2 Ablation Study

Independent and combined effects of the two branches. To examine the individual effects and combined benefits of the two branches, Table 1 compares E-D3DGS, Ablation (S) with supervision weighting alone, Ablation (T) with sparse temporal gating alone, and SAGE-GS with both branches enabled. Here, S includes learned weighting, the dynamic-region prior, soft event density, and amplitude reweighting, while T modulates per-Gaussian temporal contributions.

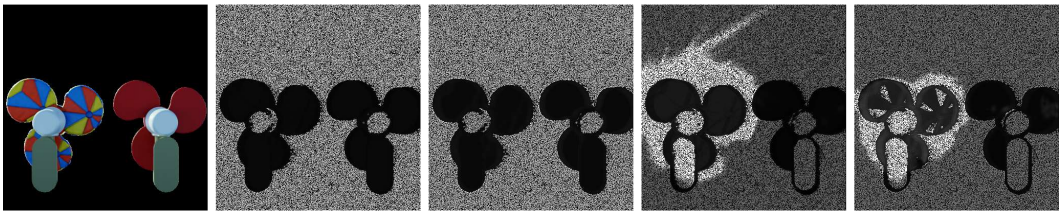
Supervision weighting accounts for most of the average performance gain, while adding temporal gating provides a smaller, scene-dependent improvement. The joint effect of the two branches depends on the scene. The full model achieves higher PSNR than either branch alone on *fan*, *celestial_globe*, *materials*, and *music_box*, whereas S-only performs better on *lego*, *hotdog*, and *man*. On *man*, for example, adding temporal gating reduces PSNR from $29.74\,\mathrm{dB}$ to $29.03\,\mathrm{dB}$. This variation may arise from the interaction between the branches during joint optimization: spatial weighting changes how individual pixels contribute to the fitting objective, while temporal gating adjusts the contributions of the primitives that render those pixels. The two mechanisms can complement each other in some scenes, but temporal suppression may also reduce primitive contributions that benefit reconstruction under the S-only configuration. Similar differences appear in real scenes. On *Jeep*, the full model reaches $27.92\,\mathrm{dB}$, exceeding S-only at $27.00\,\mathrm{dB}$ and T-only at $27.71\,\mathrm{dB}$. On *Excavator*, the full model and S-only both achieve $29.72\,\mathrm{dB}$, but the full model has higher SSIM and lower LPIPS. Overall, adding temporal gating to supervision weighting provides a small average benefit, with its effect varying across scenes and evaluation metrics.

Supervision-weighting constraints and full-image reconstruction. To distinguish the roles of the auxiliary factors and learned weighting, we compare six supervision configurations on *fan*, keeping temporal gating and the other training settings fixed. U reuses Ablation (T) with unweighted event supervision. A uses $w_{\mathrm{learn}} w_{\mathrm{conf}}$; B and C additionally incorporate the event-response factor w_{density} and dynamic prior M_d, respectively; D includes all four factors. F removes only learned weighting from D by setting $w_{\mathrm{learn}} = 1$. Table 2 reports PSNR and the corresponding SSIM and LPIPS at each configuration's best-PSNR checkpoint. Figure 2 shows relative supervision weights on a fixed sample from the independently trained configurations A–D. Each final weight map is divided by its spatial mean to compare the spatial allocation in the normalized event loss.

Table 2: Full-image reconstruction results of the supervision-weighting ablation on the *fan* scene.

	Supervision factors				Reconstruction quality		
Config.	w_{learn}	w_{density}	M_d	w_{conf}	PSNR↑	SSIM↑	LPIPS↓
U	–	–	–	–	30.36	0.963	0.025
A	✓	–	–	✓	29.41	0.959	0.026
B	✓	✓	–	✓	29.68	0.960	0.026
C	✓	–	✓	✓	29.89	0.964	0.0257
D	✓	✓	✓	✓	**30.58**	**0.966**	**0.023**
F	–	✓	✓	✓	29.93	0.962	0.0273

On *fan*, adding the event-response factor or dynamic prior to A increases PSNR by 0.27 and $0.48\,\mathrm{dB}$, respectively. D, which uses both, reaches $30.58\,\mathrm{dB}$ and exceeds B and C by 0.90 and $0.69\,\mathrm{dB}$. These comparisons favor the combined configuration on the evaluated *fan* scene.



RGB reference A B C D

Figure 2: Relative event-supervision weights on *fan*, each divided by its spatial mean and shown on a shared color scale. A–D are independently trained configurations from Table 2; the left panel is the RGB reference.

Figure 2 illustrates the spatial weight distributions associated with different supervision factors. The scattered background weights in A and B show that high weights are not concentrated on the moving objects in these configurations; background responses persist after introducing the local event-response factor. C assigns high weights to a broad region near the left fan, associating the dynamic prior with more object-focused supervision, although the allocation remains coarse around

individual blades. In contrast, D produces a more compact high-weight region and clearer blade structure, suggesting that combining the event-response factor with the dynamic prior can help refine supervision within the object region. This spatial pattern is consistent with D achieving the highest PSNR in the *fan* ablation: the complete configuration emphasizes relevant regions while retaining structural differences within them.

Removing learned weighting from D gives F, reducing PSNR by $0.65\,\mathrm{dB}$ and worsening SSIM and LPIPS. D also exceeds unweighted supervision U by $0.22\,\mathrm{dB}$, whereas A and F both fall below U. Learned weighting therefore helps within the complete configuration on *fan*; weighting alone is not sufficient. The amplitude factor is shared by configurations A–D and F. The SSIM and LPIPS results also favor D over F: SSIM rises from 0.962 to 0.966, while LPIPS falls from 0.0273 to 0.023. Relative to U, D improves SSIM from 0.963 to 0.966 and lowers LPIPS from 0.025 to 0.023. The advantage of D on this scene therefore appears across all three reported reconstruction metrics.

4.3 Analysis of Gating Sparsity

Scene-dependent gate strength. To characterize how temporal gating is distributed across scenes, we measure the near-zero gate ratio, mean gate strength, and 90th percentile for dynamic Gaussians in eight synthetic scenes at a common training endpoint of 50,000 iterations. Gaussians with $\alpha_i \leq 10^{-8}$ are counted as having near-zero gates, for which temporal suppression is considered numerically inactive. Table 3 reports these statistics.

Table 3: Gate statistics on eight synthetic scenes at 50,000 iterations. The near-zero ratio is computed using $\alpha_i \leq 10^{-8}$, and Q_{90} denotes the 90th percentile of the gate strength.

Scene	fan	lego	celestial_globe	hotdog	man	materials	waterwheel	music_box
Near-zero ratio	30.76%	24.17%	60.82%	34.56%	20.56%	46.07%	40.32%	36.91%
Mean strength	0.1023	0.0690	0.1399	0.0859	0.1937	0.0436	0.1408	0.1384
Q_{90}	0.3596	0.2462	0.5918	0.2774	0.9709	0.0464	0.5013	0.4840

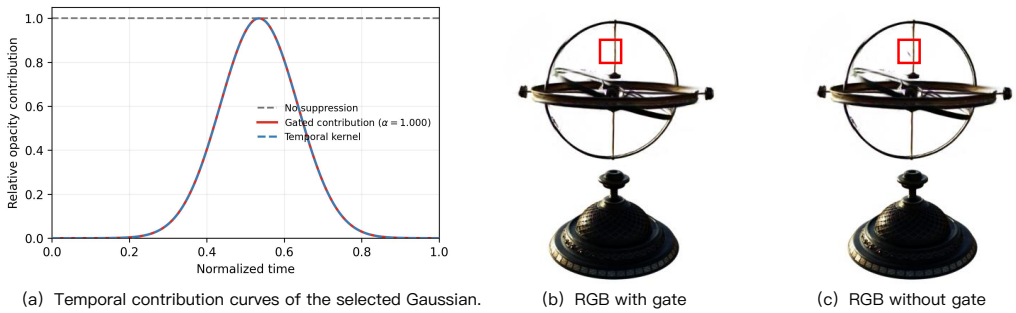
The gate statistics indicate scene-dependent, selective suppression rather than uniform temporal decay across all dynamic Gaussians. In *celestial_globe*, 60.82% of gates are near zero, yet the 90th-percentile strength reaches 0.5918. Most primitives therefore retain their base opacity, while a subset receives stronger temporal modulation. In *materials*, the 90th percentile is only 0.0464. By the gate definition, approximately 90% of primitives retain at least 95.36% of their base opacity throughout time, leaving geometric deformation to account for most temporal variation in their contributions. In contrast, *man* has the lowest near-zero ratio and a 90th percentile close to one, indicating that temporal modulation affects more primitives and concentrates some contributions around their respective temporal centers. These differences illustrate the role of per-Gaussian gating: accommodating persistent and temporally localized contributions within the same representation, with their allocation varying across scenes.

Temporal gating behavior and rendering impact. We also disable the gate of one strongly gated Gaussian by setting its contribution factor to one. All geometry, other parameters, camera settings, and rendering time remain fixed. Figure 3 compares the renderings and plots the selected primitive's kernel and contribution curve.

The selected strength is close to one, so $h_i(t)$ nearly follows $g_i(t)$. At the displayed time, away from the temporal center, disabling the gate adds a local response in the red box. This intervention verifies that the gate changes the rendered image without changing geometry. Because only one contribution factor changes, the local difference is attributable to the selected gate at this time and viewpoint. Setting the factor to one restores that primitive's base opacity while preserving its deformed position, rotation, and scale.

4.4 Qualitative Comparison

Figure 4 presents corresponding test-view renderings from E-D3DGS and SAGE-GS alongside the ground truth for three scenes. In these selected views, SAGE-GS reduces stray color responses near the rods and rings in *celestial_globe* and retains clearer cube-color boundaries in *Excavator*. Both



(a) Temporal contribution curves of the selected Gaussian. (b) RGB with gate (c) RGB without gate

Figure 3: Single-Gaussian gate intervention. (a) Learned contribution, temporal kernel, and unit baseline. (b) Learned-gate rendering. (c) Only the selected gate is disabled. Geometry, camera, and time are fixed; the red box marks the local change.

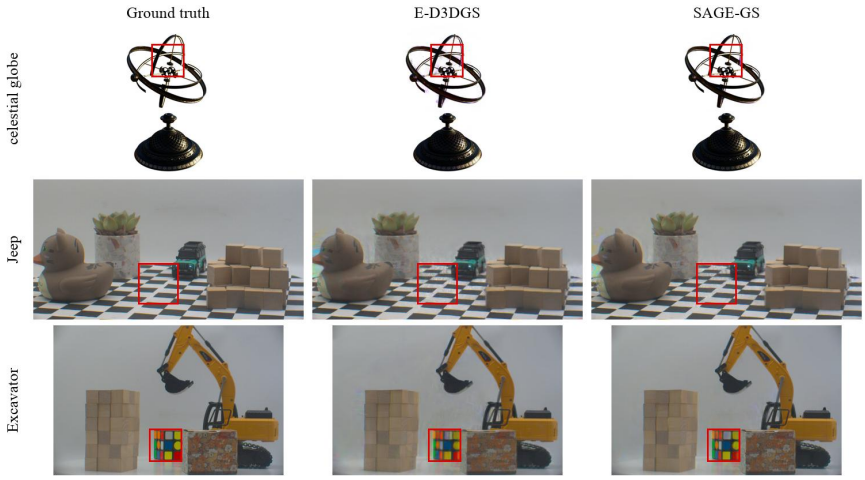


Figure 4: Qualitative comparison on *celestial_globe*, *Jeep*, and *Excavator*.

methods distort the checkerboard texture in *Jeep*. The remaining checkerboard distortion shows that these additions do not resolve all fine-texture errors. The views complement full-image metrics by locating visible differences, whereas the metrics aggregate errors across the entire image. The examples compare the local rendering behavior of the complete models.

5 Conclusion

We presented SAGE-GS, an event-guided dynamic Gaussian reconstruction method that combines spatial event-supervision weighting with sparse temporal opacity gating. The supervision branch integrates RGB–event features and auxiliary priors to regulate the relative influence of event observations across image regions, while the temporal branch selectively modulates each Gaussian's rendering contribution alongside geometric deformation. These mechanisms act on observation supervision and scene representation, respectively, allowing reconstruction to adjust both spatial supervision allocation and temporal contributions. Experiments on eight synthetic and four real scenes show that SAGE-GS improves mean PSNR over E-D3DGS by $0.52\,\mathrm{dB}$ and $0.24\,\mathrm{dB}$, respectively, while also reducing mean LPIPS. Factor ablations on *fan* further show that the complete supervision-weighting configuration outperforms unweighted supervision in PSNR, SSIM, and LPIPS, while the gating analysis demonstrates the coexistence of persistent and temporally localized contributions within the same representation. Together, these results support the effectiveness of combining spatial supervision allocation with selective temporal modulation for RGB–event dynamic reconstruction.

AI USE STATEMENT

OpenAI Codex assisted with manuscript drafting, language editing, LaTeX preparation, literature-metadata checks, and suggestions on experimental presentation and claim–evidence alignment. No experimental measurements or images were generated or altered during this revision. The authors are responsible for the manuscript and its scientific content.

REFERENCES

Xiaoyang Bai, Zhenyang Li, Weiwei Xu, Edmund Y. Lam, and Yifan Peng. ERF-GS: Reconstructing fast motion from disjoint event-rgb viewpoints. *arXiv preprint arXiv:2608.08531*, 2026. URL https://arxiv.org/abs/2608.08531.

Anish Bhattacharya, Ratnesh Madaan, Fernando Cladera, Sai Vemprala, Rogerio Bonatti, Kostas Daniilidis, Ashish Kapoor, Vijay Kumar, Nikolai Matni, and Jayesh K. Gupta. EvDNeRF: Reconstructing event data with dynamic neural radiance fields. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5846–5855, 2024. URL https://www.anishbhattacharya.com/research/evdnerf.

Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):154–180, 2022. doi: 10.1109/TPAMI.2020.3008413. URL https://doi.org/10.1109/TPAMI.2020.3008413.

Haiqian Han, Jianing Li, Henglu Wei, and Xiangyang Ji. Event-3DGS: Event-based 3D reconstruction using 3D Gaussian Splatting. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-4069. URL https://papers.nips.cc/paper_files/paper/2024/hash/e73ad1f690542144ce354637bb913c35-Abstract-Conference.html.

Inwoo Hwang, Junho Kim, and Young Min Kim. Ev-NeRF: Event based neural radiance field. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 837–847, 2023. URL https://openaccess.thecvf.com/content/WACV2023/html/Hwang_Ev-NeRF_Event_Based_Neural_Radiance_Field_WACV_2023_paper.html.

Simon Klenk, Lukas Koestler, Davide Scaramuzza, and Daniel Cremers. E-NeRF: Neural radiance fields from a moving event camera. *IEEE Robotics and Automation Letters*, 8, 2023. URL https://cvg.cit.tum.de/research/enerf.

Lucas Yunkyu Lee, Soonho Kim, Youngwook Kim, Sangmin Kim, and Jaesik Park. FreeTimeGS++: Secrets of dynamic gaussian splatting and their principles. arXiv preprint arXiv:2605.03337, 2026. URL https://arxiv.org/abs/2605.03337.

Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime gaussian feature splatting for real-time dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8508–8520, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Li_Spacetime_Gaussian_Feature_Splatting_for_Real-Time_Dynamic_View_Synthesis_CVPR_2024_paper.html.

Yunshan Qi, Lin Zhu, Yu Zhang, and Jia Li. E2NeRF: Event enhanced neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 13254–13264, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Qi_E2NeRF_Event_Enhanced_Neural_Radiance_Fields_from_Blurry_Images_ICCV_2023_paper.html.

Viktor Rudnev, Mohamed Elgharib, Christian Theobalt, and Vladislav Golyanik. EventNeRF: Neural radiance fields from a single colour event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4992–5002, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Rudnev_EventNeRF_Neural_Radiance_Fields_From_a_Single_Colour_Event_Camera_CVPR_2023_paper.html.

Yifan Wang, Peishan Yang, Zhen Xu, Jiaming Sun, Zhanhua Zhang, Yong Chen, Hujun Bao, Sida Peng, and Xiaowei Zhou. FreeTimeGS: Free gaussian primitives at anytime anywhere for dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. URL https://zju3dv.github.io/freetimegs/.

Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4D Gaussian Splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20310–20320, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wu_4D_Gaussian_Splatting_for_Real-Time_Dynamic_Scene_Rendering_CVPR_2024_paper.html.

Tianyi Xiong, Jiayi Wu, Botao He, Cornelia Fermuller, Yiannis Aloimonos, Heng Huang, and Christopher Metzler. Event3DGS: Event-based 3D Gaussian Splatting for high-speed robot egomotion. In *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 4100–4118. PMLR, 2025. URL https://proceedings.mlr.press/v270/xiong25b.html.

Wenhao Xu, Wenming Weng, Yueyi Zhang, Ruikang Xu, and Zhiwei Xiong. Event-boosted deformable 3D Gaussians for dynamic scene reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 28334–28343, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Xu_Event-boosted_Deformable_3D_Gaussians_for_Dynamic_Scene_Reconstruction_ICCV_2025_paper.html.

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20331–20341, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Yang_Deformable_3D_Gaussians_for_High-Fidelity_Monocular_Dynamic_Scene_Reconstruction_CVPR_2024_paper.html.

Xiaoting Yin, Hao Shi, Yuhan Bao, Zhenshan Bing, Yiyi Liao, Kailun Yang, and Kaiwei Wang. E-3DGS: 3D gaussian splatting with exposure and motion events. *Applied Optics*, 64(14):3897–3908, 2025. doi: 10.1364/AO.557565. URL https://opg.optica.org/ao/abstract.cfm?uri=ao-64-14-3897.

Mingchen Zhong, Xin Lu, Dong Li, Senyan Xu, Ruixuan Jiang, Xueyang Fu, and Baocai Yin. CompEvent: Complex-valued Event-RGB fusion for low-light video enhancement and deblurring. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(16):13530–13538, 2026. doi: 10.1609/aaai.v40i16.38358. URL https://ojs.aaai.org/index.php/AAAI/article/view/38358.