PHYSMOKE: PHYSICAL SURFACE AND VOLUME GAUSSIANS FOR CLEAR RECONSTRUCTION THROUGH SMOKE

Anonymous authors
Paper under double-blind review

ABSTRACT

Recovering a clear scene from multiview observations of moving smoke is difficult because visible images entangle persistent surfaces with a transient participating medium. Thermal observations provide complementary structural cues, but persistent surfaces and dynamic smoke have different physical supports that call for different representations. We introduce PhySmoke, a framework guided by physics that models the scene with two-dimensional (2D) Gaussian surfaces constrained by thermal observations and deformable three-dimensional (3D) Gaussian smoke. A unified renderer composites both representations along each viewing ray, allowing their interaction to be learned consistently. To recover appearance lost during surface and smoke separation, we further use the estimated smoke contribution to guide the restoration of visible detail. Experiments on scenes with smoke show substantial improvements in the reconstruction of clear scenes across image quality metrics, together with a more compact surface representation. These results highlight the value of matching scene representations and detail recovery to the physical structure of observations through smoke.

1 INTRODUCTION

Recovering a clear scene from multiview observations of moving smoke is important for inspection, emergency response, and robotic perception. The visible red-green-blue (RGB) signal mixes persistent surface appearance with attenuation and scattered light that change over time, making it difficult to tell whether a weak image feature belongs to the scene or the medium. Multiple views provide geometric constraints, while thermal observations offer complementary structural evidence through smoke. The central challenge is to use these cues to separate the stable scene from the dynamic medium and recover its clear appearance.

This separation is also a representation problem. Existing multimodal Gaussian approaches can decompose observations into static and dynamic components (Jain et al., 2025), but representing both as volumes does not reflect their different physical supports. Solid scene radiance arises primarily at surfaces; smoke occupies a spatial volume and changes over time. Moreover, the two components interact along each viewing ray. Their relative depth determines visibility and therefore cannot be captured by simply blending independently rendered images.

We propose PhySmoke, a surface and volume representation grounded in physical structure for clear reconstruction through smoke. Thermal observations constrain the persistent scene as native 2D Gaussian surfels (Huang et al., 2024), while deformable 3D Gaussians describe dynamic smoke. A mixed rasterizer projects each primitive with its native opacity, orders surfaces and smoke together by depth, and composites them under a shared transmittance model. This lets visible and thermal evidence constrain the components they can explain while preserving their interaction during rendering and optimization.

Separating surface and smoke can suppress genuine appearance detail along with the medium. Recovering detail directly from the smoky observation is therefore useful only where the estimated medium contribution is small. We compare renders of the full scene with renders of the surface alone to estimate that contribution and use an exponential reliability rule to control detail recovery. An independent volumetric reconstruction provides complementary appearance evidence for fusion at

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 scene level. Together, these steps use the learned decomposition both to reconstruct the scene and to decide which visible details can be restored.

Our contributions are:

- A multimodal representation that matches persistent surfaces with 2D Gaussian surfels and dynamic smoke with deformable 3D Gaussians.
- A mixed rasterizer that jointly orders and composites both primitive types along viewing rays while retaining their native parameterizations.
- A detail recovery mechanism guided by reliability estimates that uses the estimated smoke contribution to restore visible detail without indiscriminately reintroducing the medium.

Experiments on scenes obscured by smoke evaluate reconstruction quality, surface compactness, and the contribution of each design choice.

2 Related Work

Image and video restoration through participating media. Classical atmospheric models describe observed radiance as attenuated scene radiance plus path radiance (Narasimhan & Nayar, 2002). Modern restoration methods use hand-crafted transmission priors (He et al., 2011; Berman et al., 2016) or learn a direct image mapping informed by the scattering model (Li et al., 2017; Chen et al., 2021; Xu et al., 2023). Real paired haze data (Ancuti et al., 2018) and physically simulated fog (Xie et al., 2024) illustrate the range of capture and evaluation settings. These methods are primarily predictors operating in image or video space. They do not maintain an explicit separation in three dimensions between a persistent scene and temporally varying smoke. Thermal imaging can preserve visibility under aerosol corruption better than visible imaging (Erlenbusch et al., 2023), while full scattering tomography requires substantially stronger capture and calibrated transport assumptions (Levis et al., 2017). We pursue a middle ground: an explicit multiview representation with a structural physical prior, without claiming to recover calibrated scattering coefficients.

Gaussian radiance fields. NeRF (Mildenhall et al., 2020) and its extensions such as Mip-NeRF 360 (Barron et al., 2022) use volumetric rendering for multiview reconstruction. Surface-focused neural fields instead constrain geometry through implicit surfaces (Wang et al., 2021; Yariv et al., 2021; Yu et al., 2022). 3D Gaussian Splatting (Kerbl et al., 2023) represents a radiance field with anisotropic volumetric primitives and a rasterizer that accounts for visibility. Dynamic radiance fields and Gaussian variants model time-dependent motion (Pumarola et al., 2021; Luiten et al., 2024; Wu et al., 2024; Yang et al., 2024b). 2D Gaussian Splatting (Huang et al., 2024) collapses each primitive onto an oriented disk and computes ray intersections with surfels under perspective projection, making it a natural model for surfaces. MAtCha Gaussians further use surfels to represent surface geometry (Guedon et al., 2025). We use neither family as a universal representation. The surface and participating medium receive different primitive geometry because their physical support differs.

Static and transient decomposition. RobustNeRF (Sabour et al., 2023), NeRF in the Wild (Martin-Brualla et al., 2021), and the approach of Ren et al. (2024) suppress or explain inconsistent observations. EmerNeRF separates static and dynamic fields in driving scenes (Yang et al., 2024a). DeSplat (Wang et al., 2025) and RobustSplat (Fu et al., 2025) separate distractors in Gaussian scenes. HybridGS (Lin et al., 2025) assigns 2D Gaussians specific to each view to transient distractors and 3D Gaussians to the static scene. Our physical assignment is different: persistent surfaces are 2D, while smoke is a shared dynamic 3D medium. More importantly, both families are interleaved before compositing rather than blended as completed layers. SmokeSeer is the closest system because it uses the same RGB and thermal smoke task; our experiments preserve its losses and dynamic smoke model to isolate representation, rendering, and recovery.

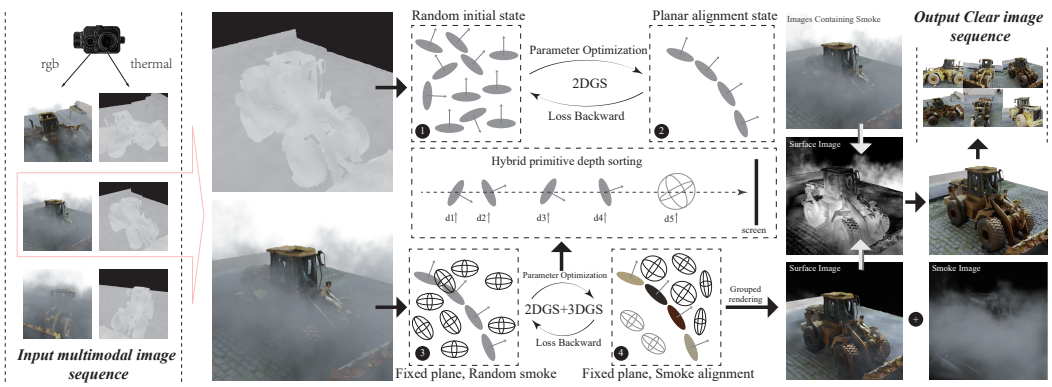


Figure 1: **Overview of PhySmoke.** Synchronized thermal and smoky RGB sequences guide reconstruction in two stages. Thermal observations first align native 2D Gaussian surfels with the persistent scene (top). The hybrid stage introduces deformable 3D Gaussians for smoke and jointly optimizes both primitive families through compositing after sorting by depth with shared transmittance (bottom). Grouped rendering separates surface and smoke contributions; the render of the surface alone provides an initial clear estimate for the detail recovery guided by reliability described in Section 3.4.

3 METHOD

Figure 1 summarizes the complete pipeline. We first use multimodal observations to build a surface and volume scene representation, then render both components with a shared visibility model. The resulting surface and joint renders guide the final recovery of clear images.

3.1 MULTIMODAL OBSERVATION MODEL

Let $\mathbf{r}_v(s) = \mathbf{o}_v + s\mathbf{d}_v$ denote a ray from observed camera v, with camera center $\mathbf{o}_v$, unit ray direction $\mathbf{d}_v$, and distance $s \geq 0$. The index t denotes capture time. The capture contains a smoky RGB image $\mathbf{I}_{v,t}$, a registered thermal image $\mathbf{H}_{v,t}$, and camera parameters Π_v. The desired output is clear visible radiance at the same camera.

For a ray intersecting a surface at distance $s_\star$, a simplified single-scattering model writes visible radiance as:

$$\mathbf{I}_{v,t} = T_t(s_\star)\,\mathbf{L}_S(\mathbf{r}_v(s_\star)) + \int_0^{s_\star} T_t(s)\,\sigma_t(\mathbf{r}_v(s))\mathbf{L}_V(\mathbf{r}_v(s), t)\,\mathrm{d}s + \boldsymbol{\epsilon}, \tag{1}$$

Here $\mathbf{L}_S$ is surface RGB radiance, $\mathbf{L}_V$ is effective medium source radiance, σ_t is effective medium density, and $\boldsymbol{\epsilon}$ absorbs observation and model error. Pixel arguments are suppressed. Transmittance along the ray is:

$$T_t(s) = \exp\left(-\int_0^{s} \sigma_t(\mathbf{r}_v(u))\,\mathrm{d}u\right) \tag{2}$$

The variable u is the integration distance along the same ray. Equation 1 motivates, but does not fully specify, our renderer: multiple scattering and material properties that depend on wavelength parameters are absorbed into learned radiance and opacity.

The thermal observation follows a different spectral response. We use the reduced influence of visible smoke as a modeling cue:

$$\mathbf{H}_{v,t} = \mathcal{P}_{\mathrm{th}}(\partial\Omega; \Pi_v) + \boldsymbol{\eta}_{v,t}, \tag{3}$$

where Ω is the persistent solid region, $\partial\Omega$ is its boundary, $\mathcal{P}_{\mathrm{th}}$ is an abstract thermal image formation operator conditioned on camera parameters Π_v, and $\boldsymbol{\eta}_{v,t}$ is unmodeled thermal error. This approximation does not assert that thermal images are unaffected by smoke; it states that they are a more stable source of boundary evidence than visible observations in this capture setting.

3.2 PHYSICAL SUPPORT DETERMINES PRIMITIVE DIMENSION

The static component is supported on a boundary, whereas the medium density is defined in volume:

$$\mathcal{S} \subset \partial\Omega, \quad \dim(\partial\Omega) = 2, \qquad \sigma : \mathbb{R}^3 \times \mathcal{T} \to \mathbb{R}_{\geq 0}. \tag{4}$$

Here $\mathcal{S}$ is the surface set, $\mathcal{T}$ is the set of capture times, and σ maps a 3D position and time to nonnegative density. We encode $\mathcal{S} = \{s_i\}$ with oriented 2D Gaussian surfels indexed by i. A surfel has center $\boldsymbol{\mu}_i$, tangent scales (a_i, b_i), orientation, opacity, and appearance. Native ray intersection with surfels produces tangent coordinates $\mathbf{u}_i(\mathbf{p})$ at image pixel $\mathbf{p}$, giving its opacity:

$$\alpha_i^S(\mathbf{p}) = \min\big(\alpha_{\max}, o_i \exp[-\tfrac{1}{2}\|\mathbf{u}_i(\mathbf{p})\|_2^2]\big)\,. \tag{5}$$

Here o_i is the surfel's opacity parameter, $\alpha_{\max}$ is the opacity cap, and α_i^S is its projected opacity. The smoke set $\mathcal{V}_t = \{g_k(t)\}$ contains deformable anisotropic 3D Gaussians. After projection, its opacity is:

$$\alpha_k^V(\mathbf{p}, t) = \min\big(\alpha_{\max}, o_k \exp[-\tfrac{1}{2}\Delta\mathbf{p}^\top \mathbf{Q}_{k,t}\Delta\mathbf{p}]\big)\,, \tag{6}$$

Here k indexes smoke Gaussians, o_k is a smoke Gaussian's opacity parameter, $\Delta\mathbf{p}$ is the pixel offset from its projected center, $\mathbf{Q}_{k,t}$ is its conic precision (inverse projected covariance), and α_k^V is its projected opacity. The superscripts S and V mark surface and volume primitives, respectively. Equation 5 concentrates capacity near a surface; Equation 6 retains the support required by a diffuse moving medium.

3.3 ONE DEPTH ORDER AND ONE TRANSMITTANCE CHAIN

For a pixel $\mathbf{p}$, the rasterizer forms one list $\mathcal{L}(\mathbf{p})$ containing all intersecting surface and smoke primitives, sorted front to back by projected center depth. A type tag selects Equation 5 or Equation 6; both families then obey:

$$T_1(\mathbf{p}) = 1, \qquad T_i(\mathbf{p}) = \prod_{j<i}(1 - \alpha_j(\mathbf{p})), \tag{7}$$

$$\mathbf{J}(\mathbf{p}) = \sum_{i \in \mathcal{L}(\mathbf{p})} T_i(\mathbf{p})\alpha_i(\mathbf{p})\mathbf{c}_i. \tag{8}$$

In this recurrence, i indexes the depth-ordered combined list, $j < i$ denotes earlier primitives, α_i selects the surface or smoke opacity, $\mathbf{c}_i$ is primitive RGB color, T_i is discrete transmittance before primitive i, and $\mathbf{J}$ is the joint RGB render. The image of the surface alone $\mathbf{S}$ is obtained by omitting $\mathcal{V}_t$. Independent blending of rendered images cannot express a ray on which smoke lies both in front of and behind a surface, and it assigns different optimization gradients. Our CUDA (Compute Unified Device Architecture) backward pass traverses the combined list in reverse and returns derivatives to the native surfel transform and volumetric conic parameters.

3.4 DETAIL RECOVERY GUIDED BY RELIABILITY

The physical representation produces a clear estimate $\mathbf{S}_h$, but decomposition error may place fine surface detail into the smoke component. Let h denote this heterogeneous representation of surfaces and volumes and b the original volumetric representation from SmokeSeer. Each produces a joint render $\mathbf{J}_r$ and a surface render $\mathbf{S}_r$, $r \in \{h, b\}$.

3.4.1 EVIDENCE FOR REPRESENTATION CHOICE AT THE SCENE LEVEL

Both representations were optimized to explain the available smoky input. Let $\mathcal{P}$ contain all evaluated scene pixels, and write $\mathbf{I}(\mathbf{p})$ for their observed smoky RGB. The norm $\|\cdot\|_1$ sums absolute RGB-channel errors. We compare the representations' mean joint reconstruction errors:

$$e_r = \frac{1}{|\mathcal{P}|}\sum_{\mathbf{p}\in\mathcal{P}} \|\mathbf{I}(\mathbf{p}) - \mathbf{J}_r(\mathbf{p})\|_1, \qquad \rho = \frac{e_b}{e_h}. \tag{9}$$

Here e_r is the error for representation r, $|\mathcal{P}|$ is the pixel count, and $\rho = e_b/e_h$ compares the volumetric baseline with the heterogeneous representation. We use a conservative, fixed rule with three levels:

$$w = g(\rho) = \begin{cases} 1, & \rho \geq 0.95, \\ 0.25, & 0.90 \leq \rho < 0.95, \\ 0.5, & \rho < 0.90, \end{cases} \tag{10}$$

The function g maps ρ to the scene-level fusion weight w, which is frozen for the scene. The final branch deliberately avoids a hard switch when both reconstructions are poor. The thresholds in Equation 10 are global constants rather than choices for each scene.

The fused surface and unexplained RGB residual are:

$$\overline{\mathbf{S}} = w\mathbf{S}_h + (1-w)\mathbf{S}_b, \tag{11}$$

$$\mathbf{E} = w(\mathbf{I} - \mathbf{J}_h) + (1-w)(\mathbf{I} - \mathbf{J}_b). \tag{12}$$

Here $\overline{\mathbf{S}}$ is the fused surface render and $\mathbf{E}$ is the fused smoky RGB reconstruction residual. This comparison across representations gives more weight to the physical hypothesis that better reproduces the observed smoky radiance.

3.4.2 Smoke contribution as a proxy for optical thickness

The difference between joint and surface renders estimates the radiance change assigned to smoke:

$$D(\mathbf{p}) = w\,\overline{|\mathbf{J}_h - \mathbf{S}_h|}_c + (1-w)\,\overline{|\mathbf{J}_b - \mathbf{S}_b|}_c, \tag{13}$$

where the bar with subscript c averages the absolute difference over RGB channels, and $D(\mathbf{p})$ is the resulting per-pixel smoke score. We do not equate D with calibrated optical thickness. We use it as an ordinal proxy: a large difference means the learned image formation model attributes more of the pixel to the medium. Motivated by Equation 2, the reliability of residual detail is:

$$C(\mathbf{p}) = \exp[-\beta D(\mathbf{p})]. \tag{14}$$

Here $C(\mathbf{p})$ is a per-pixel reliability weight and $\beta > 0$ controls how quickly it falls as D grows. Our recovered clear image is:

$$\widehat{\mathbf{I}}(\mathbf{p}) = \mathrm{clip}\big(\overline{\mathbf{S}}(\mathbf{p}) + \lambda C(\mathbf{p})\mathbf{E}(\mathbf{p}), 0, 1\big), \tag{15}$$

The gain λ scales residual detail, and clip clamps each output channel to $[0, 1]$. We use $\lambda = 2$ and $\beta = 8$.

Equation 15 has two useful boundary behaviors. If the models predict negligible smoke, $D \to 0$ and $C \to 1$, so unmodeled RGB detail is recovered. If the predicted medium effect grows, $C \to 0$ and the result returns to the clear surface estimate. More generally, before clipping, the following bound holds:

$$\|\widehat{\mathbf{I}} - \overline{\mathbf{S}}\|_\infty \le \lambda e^{-\beta D}\|\mathbf{E}\|_\infty \tag{16}$$

This is a per-pixel bound: $\|\cdot\|_\infty$ is the largest absolute RGB channel value and D denotes $D(\mathbf{p})$ at that pixel. Thus the maximum contamination admitted from the smoky residual decreases monotonically with the modeled medium contribution. This bound explains why the gate can preserve the reconstruction benefit of residual recovery without inheriting its full perceptual error.

4 Experiments

4.1 Data, task, and implementation

We evaluate on the public synthetic benchmark released with SmokeSeer (Jain et al., 2025). It was generated with Blender's Mantaflow smoke simulator and includes both individual objects and larger environments. Our evaluation uses nine scenes and 1,055 posed multiview observations (110 to 150 per scene), with smoky RGB and corresponding thermal images. Clear RGB images provide references for quantitative evaluation. Unless stated otherwise, all experiments in this section use the NoDense protocol with the released initialization of 100,000 random points. Section 4.5 additionally evaluates Dense, which uses COLMAP multiview stereo (MVS) points aligned to the same scenes while keeping the recovery settings selected on NoDense unchanged.

The task is clear reconstruction at observed input views. The method may use the smoky RGB observation at that pose, so this is not a claim about synthesizing novel views. We report peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) (Wang et al., 2004), and learned perceptual image patch similarity (LPIPS) with Visual Geometry Group (VGG) network features (Zhang et al.,

Table 1: Clear reconstruction at observed views and compactness of the clear surface on nine scenes. Image metrics are macro averages. The surface ratio is the mean of each scene's native 2D primitive count after the first stage divided by its matched 3D surface count; the physical and complete methods share this surface. The ratio does not measure total system primitives. N/A indicates that the baseline has no corresponding 2D surface.

Method	PSNR ↑	SSIM ↑	LPIPS ↓	Surface ratio ↓
SmokeSeer	20.61	0.7569	0.1869	N/A
Physical representation	21.08	0.8029	0.1935	0.476
Complete method	**21.61**	**0.8141**	**0.1583**	0.476

Table 2: Recovery ablation relative to the physical surface and volume model. Δ denotes the variant minus that model; positive PSNR/SSIM and negative LPIPS are better. PSNR wins counts scenes with positive ΔPSNR.

Variant	ΔPSNR	ΔSSIM	ΔLPIPS	PSNR wins
Scene fusion	+0.208	+0.0039	-0.0088	2/9
Ungated residual recovery	+0.527	-0.0667	+0.0201	9/9
Selective recovery	**+0.528**	**+0.0112**	**-0.0351**	**9/9**

2018). We average over views within a scene and then average the nine scene means. Higher PSNR and SSIM and lower LPIPS are better; dB denotes decibels.

The first stage optimizes the surface using thermal observations for 15,000 iterations. The second stage adds the dynamic smoke set and uses 15,000 joint forward and backward calls. The losses, seed, schedule, and smoke deformation network follow SmokeSeer. For native 2D surfels, the weights for normal consistency and distortion are 0.05 and 0.01, activated at iterations 7,000 and 3,000. Recovery introduces no training. We choose $(\lambda, \beta) = (2, 8)$ on NoDense and freeze it for Dense. Statistical analysis is detailed in Appendix D.

4.2 Overall clear reconstruction

Across the benchmark, the complete method achieves the best macro average on all three metrics (Table 1): 21.61 dB PSNR, 0.8141 SSIM, and 0.1583 LPIPS. Relative to SmokeSeer, these correspond to gains of 0.995 dB and 0.0572 in PSNR and SSIM, and a 0.0286 reduction in LPIPS. The physical surface and volume representation alone raises PSNR by 0.467 dB and SSIM by 0.0460, while its LPIPS is 0.0065 worse. Recovery guided by reliability then improves all three macro metrics, showing the benefit of restoring appearance after separating surfaces from smoke.

The clear surface representation is also more compact. Across the same nine scenes, its native 2D surfels use $0.476\times$ as many primitives as a matched 3D Gaussian surface on average, a 52.4% reduction. This count compares the surface components after the first stage; Section 4.6 reports the counts for each scene and thermal reconstruction quality.

Relative to the physical surface and volume model, detail recovery adds +0.528 dB in PSNR, raises SSIM by +0.0112, and lowers LPIPS by 0.0351 on average. PSNR improves in all nine scenes, while LPIPS decreases in five. Results for each scene and statistical tests are reported in Appendices C and D.

Figure 2 compares matched views from Hotdog, Caterpillar, and Mars, whose gains in mean PSNR over SmokeSeer are $+2.22$, $+2.16$, and $+1.50$ dB, respectively. The fixed selection rule within each scene avoids choosing the frame with the largest gain. Table 4 reports PSNR for every scene.

4.3 Reliable detail recovery improves reconstruction

Fusion at the scene level alone improves every macro metric, but its PSNR gain is concentrated in two scenes. Ungated residual recovery improves PSNR by 0.527 dB in 9/9 scenes, establishing

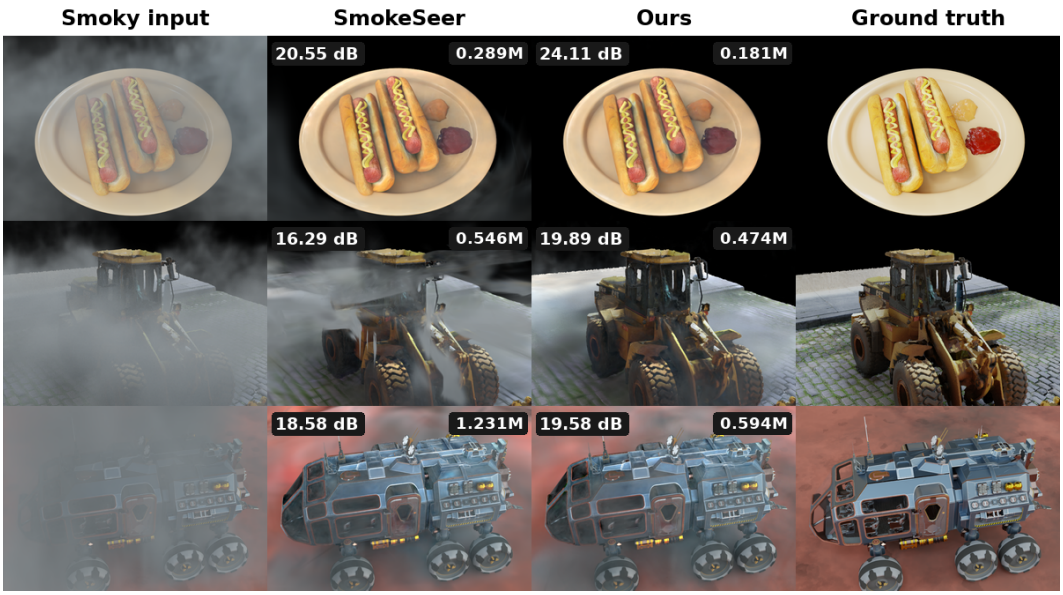


Figure 2: **Clear reconstruction on the three scenes with the largest gains in mean PSNR over SmokeSeer.** Columns show the smoky RGB input, SmokeSeer's clear prediction, our complete method, and the clear reference (evaluation only). Rows show Hotdog, Caterpillar, and Mars from top to bottom. Labels at the upper left show PSNR for the selected view; labels at the upper right of the two prediction columns show final surface primitive counts in millions (M). For the complete method, this count refers to its native 2D surface component. Each row uses one camera pose.

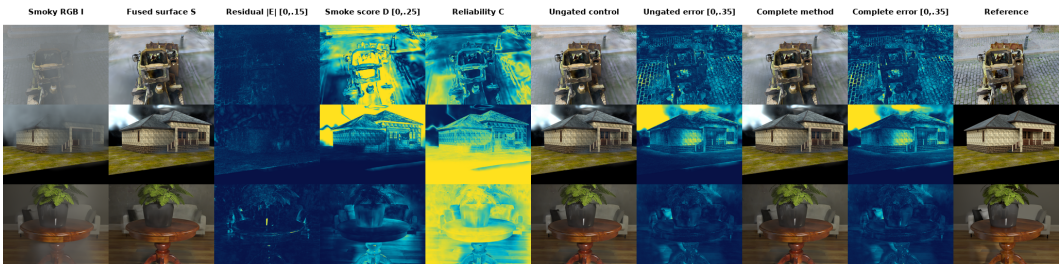


Figure 3: **Evidence for reliable detail recovery at the pixel level.** The difference between joint and surface renders yields smoke score D and reliability $C = \exp(-8D)$; low reliability suppresses residual transfer where the modeled smoke effect is strong. Ungated and complete error maps share the $[0, 0.35]$ scale (blue: low; yellow: high).

that the joint reconstruction omits useful detail. It simultaneously reduces SSIM by 0.0667 and worsens LPIPS by 0.0201, showing that the same residual contains smoke or representation error. The exponential reliability retains the PSNR gain and reverses both secondary metrics. This controlled comparison is the direct evidence for Equation 14; an unqualified residual skip is not sufficient.

Figure 3 shows how the gate preserves useful detail while reducing the smoky structure transferred by ungated recovery.

Applying the same recovery mechanism to SmokeSeer alone gives $+0.282$ dB PSNR, $+0.0075$ SSIM, and -0.0192 LPIPS relative to SmokeSeer, with PSNR improving in 9/9 scenes. Its larger gain on the physical representation ($+0.528$ dB) supports complementarity: the surface and volume model provides a cleaner initial separation and a more informative medium score, while the volumetric model supplies alternative appearance evidence in difficult scenes.

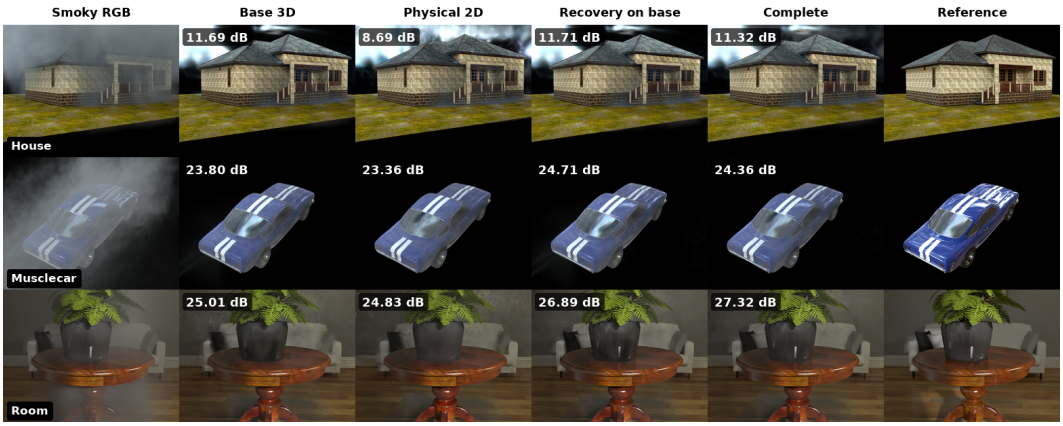


Figure 4: **Complementary effects of physical representation and detail recovery.** The three difficult scenes illustrate cases where the physical model alone loses visible detail. Recovery on SmokeSeer and the complete method use the same recovery rule. PSNR labels refer to the displayed views; the reference is used only for evaluation. The six scene comparison is in Appendix B.

Figure 4 shows the missing appearance detail in House, Musclecar, and Room. Recovery restores detail on either representation, while the complete method combines that detail with physical surface and smoke separation.

Attenuation sensitivity. Weak attenuation at $\beta = 4$ admits more residual and attains the largest PSNR gain, but decreases SSIM. Strong attenuation at $\beta = 16$ is more conservative and improves SSIM, at the cost of PSNR. The selected $\beta = 8$ retains more than 0.5 dB while improving both secondary metrics, and gives the largest LPIPS reduction among the three settings. All settings improve PSNR in every scene, so the effect is not isolated to one attenuation constant. The full sensitivity results are in Appendix E.

4.4 Joint transport of primitives improves reconstruction

We isolate the renderer using the same native 2D surface, 3D smoke checkpoint, losses, and optimization schedule. The control independently renders the two families and orders the completed layers by expected depth. Sorting individual primitives and sharing transmittance instead improves PSNR by +1.17 dB, SSIM by +0.0225, and LPIPS by -0.0154 over all nine scenes. The exact PSNR sign flip test yields a p-value of 0.0039. Forward passes containing only 2D or only 3D primitives match their native renderers; parameter gradients agree within 3×10^{-5}, and 100 randomized mixed forward/backward graphs contain finite values. This isolates the gain from transport ordering rather than from a changed representation.



Figure 5: **Rendering individual primitives versus complete layers.** With identical surface and smoke checkpoints, joint depth ordering and shared transmittance reduce a bright compositing artifact while preserving the object boundary.

The corresponding absolute-error maps appear in Appendix B (Figure 7).

4.5 Transfer to Dense initialization

Table 3: Absolute macro metrics under the primary NoDense initialization and the Dense MVS initialization. Recovery parameters selected on NoDense are transferred unchanged.

	NoDense			Dense transfer		
Method	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
SmokeSeer	20.61	0.7569	0.1869	21.53	0.8396	0.1549
Physical representation	21.08	0.8029	0.1935	21.41	0.8272	0.1662
Complete method	21.61	0.8141	0.1583	21.68	0.8344	0.1453

On Dense, recovery raises the physical model's PSNR by 0.270 dB on average, with gains in eight of the nine scenes. It raises SSIM by 0.0072 and lowers LPIPS by 0.0209. The complete model also exceeds the Dense SmokeSeer in PSNR ($+0.149$ dB) and LPIPS (-0.0097), although SSIM remains 0.0053 lower. Dense contains the same scene identities, so this experiment tests transfer of the recovery settings across initialization protocols. The statistical analysis is reported in Appendix D.

4.6 Surface compactness

The native 2D surface uses between 0.407 and 0.562 times as many primitives as its matched 3D Gaussian surface across the nine scenes, with mean $0.476\times$. The mean thermal PSNR difference after the first stage is $+0.65$ dB but varies from -1.67 to $+8.39$ dB. Primitive count is therefore an efficiency proxy rather than evidence of equal fidelity. Table 7 in Appendix F reports the values for each scene.

5 Limitations

Our model captures surface support and monotone attenuation, not calibrated scattering or thermal transport. The reliability score and the thresholds applied to each scene are empirical. Evaluation uses one seed on nine synthetic scene identities at observed views; the Dense protocol changes initialization, not scene content, and residual recovery requires a smoky observation at the target pose. The complete method does not win on every scene and metric. Real paired RGB and thermal data, previously unseen scenes, and repeated seeds are needed before deployment claims.

6 Conclusion

We presented an approach guided by physics to clear reconstruction through dynamic smoke. Thermal cues anchor a native 2D surface, while smoky RGB constrains appearance and a deformable 3D medium. A shared transmittance chain renders both primitive families. The modeled smoke contribution then controls the recovery of RGB detail lost during surface and smoke separation. Together, representation and recovery improve all three macro metrics and add more than 0.5 dB over the physical representation in the primary evaluation.

Reproducibility statement

The supplement provides frozen predictions, weights for each scene, metric outputs, generation scripts, and the bootstrap seed. Equations 10 to 15 specify the recovery parameters. We plan to release the source code, model weights, predictions, and evaluation scripts upon acceptance.

AI use statement

Generative AI assisted language editing, experiment organization, and draft analysis scripts. The authors verified the numerical claims and take responsibility for the paper and artifacts.

REFERENCES

Codruta O. Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-HAZE: A dehazing benchmark with real hazy and haze-free outdoor images. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 754–762, 2018.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5470–5479, 2022.

Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1674–1682, 2016.

Zeyuan Chen, Yangchao Wang, Yang Yang, and Dong Liu. PSD: Principled synthetic-to-real dehazing guided by physical priors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7180–7189, 2021.

Daniel Erlenbusch, Maxim K"ammerer, Laura L"offler, and Annette Stahl. Thermal infrared single image dehazing and blind image quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023.

Chuanyu Fu, Yuqi Zhang, Kunbin Yao, Guanying Chen, Yuan Xiong, Chuan Huang, Shuguang Cui, and Xiaochun Cao. RobustSplat: Decoupling densification and dynamics for transient-free 3dgs. In *IEEE/CVF International Conference on Computer Vision*, pp. 27126–27136, 2025.

Antoine Guedon, Tomoki Ichikawa, Kohei Yamashita, and Ko Nishino. MAtCha gaussians: Atlas of charts for high-quality geometry and photorealism from sparse views. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6001–6011, 2025.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. doi: 10.1109/TPAMI.2010.168.

Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *ACM SIGGRAPH Conference Papers*, 2024.

Neham Jain, Andrew Jong, Sebastian Scherer, and Ioannis Gkioulekas. SmokeSeer: 3d gaussian splatting for smoke removal and scene reconstruction. *arXiv preprint arXiv:2509.17329*, 2025.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023.

Aviad Levis, Yoav Y. Schechner, Amit Aides, and Anthony B. Davis. Multiple-scattering microphysics tomography. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision*, 2017.

Jingyu Lin, Jiaqi Gu, Lubin Fan, Bojian Wu, Yujing Lou, Renjie Chen, Ligang Liu, and Jieping Ye. HybridGS: Decoupling transients and statics with 2d and 3d gaussian splatting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 788–797, 2025.

Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *International Conference on 3D Vision*, 2024. doi: 10.1109/3DV62453.2024.00044.

Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF in the Wild: Neural radiance fields for unconstrained photo collections. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, 2020.

Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, 2002.

Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural radiance fields for dynamic scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10318–10327, 2021.

Weining Ren, Zihan Zhu, Boyang Sun, Jiaqi Chen, Marc Pollefeys, and Songyou Peng. NeRF On-the-go: Exploiting uncertainty for distractor-free nerfs in the wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8931–8940, 2024.

Sara Sabour, Suhani Vora, Daniel Duckworth, Ivan Krasin, David J. Fleet, and Andrea Tagliasacchi. RobustNeRF: Ignoring distractors with robust losses. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20626–20636, 2023.

Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Yihao Wang, Marcus Klasson, Matias Turkulainen, Shuzhe Wang, Juho Kannala, and Arno Solin. DeSplat: Decomposed gaussian splatting for distractor-free rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 722–732, 2025.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Yiming Xie, Henglu Wei, Zhenyi Liu, Xiaoyu Wang, and Xiangyang Ji. SynFog: A photo-realistic synthetic fog dataset based on end-to-end imaging simulation for advancing real-world defogging in autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21763–21772, 2024.

Jiaqi Xu, Xiaowei Hu, Lei Zhu, Qi Dou, Jifeng Dai, Yu Qiao, and Pheng-Ann Heng. Video dehazing via a multi-range temporal alignment network with physical prior. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18053–18062, 2023.

Jiawei Yang, Boris Ivanovic, Or Litany, Xinshuo Weng, Seung Wook Kim, Boyi Li, Tong Che, Danfei Xu, Sanja Fidler, Marco Pavone, and Yue Wang. EmerNeRF: Emergent spatial-temporal scene decomposition via self-supervision. In *International Conference on Learning Representations*, 2024a.

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20331–20341, 2024b.

Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Zehao Yu, Songyou Peng, Michael Niemeyer, Torsten Sattler, and Andreas Geiger. MonoSDF: Exploring monocular geometric cues for neural implicit surface reconstruction. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.