VEILSPLAT: SEPARATING CHANGING SMOKE FROM 3D GAUSSIAN SCENES

Anonymous authors
Paper under double-blind review

ABSTRACT

Smoke-free novel-view synthesis seeks to recover a clear, consistent scene from images captured through changing smoke. Existing approaches often rely on pretrained depth models or generative restoration priors. We introduce VeilSplat, a physics-based Gaussian reconstruction method that jointly recovers scene and medium from posed smoky RGB images without external pretrained networks. Our key principle is to separate surface coverage from medium transmission: smoke attenuates and redistributes radiance while the direct rendering branch retains the scene's occlusion weights. A compact density field combines shared spatial bases with per-view coefficients to capture spatially nonuniform smoke variation. Path-sampled optical thickness couples this field to attenuated direct radiance, broadened forward scattering, and airlight. The recovered scene supports clean novel views through ordinary Gaussian splatting. On five RealX3D smoke scenes, VeilSplat leads all five evaluated baselines with mean PSNR/SSIM/LPIPS of 16.30/0.6288/0.5083, gaining 5.44 dB PSNR over the strongest baseline. Matched component studies establish the benefits of coverage separation, capture-varying transmission and forward scattering.

1 INTRODUCTION

A smoky photograph records objects through an intervening, changing volume. Reconstructing the objects as a clear 3D scene would make the captured environment inspectable from new viewpoints, with its colors and surface detail separated from the smoke present during capture. This is a multi-view recovery task: a single scene must explain several obscured observations and support a camera that was not used to fit it.

Ordinary Gaussian splatting learns geometry and appearance that reproduce the training images (Kerbl et al., 2023). Through smoke, matching those images can place the veil into Gaussian color, scale or opacity. A changing obstruction then becomes a persistent property of the reconstructed object. The central design question is where to assign this variation. Surface coverage determines which object is visible; medium transmission determines how its radiance reaches the camera. Recovering a clear scene calls for separate representations of these two roles.

Scattering-aware reconstruction addresses this ambiguity by modeling how scene radiance reaches the sensor. SeaThru-NeRF separates object radiance and medium effects (Levy et al., 2023); WaterSplatting and SeaSplat bring such decompositions to Gaussian reconstruction (Li et al., 2025; Yang et al., 2025). Other approaches incorporate pretrained depth priors, as in DehazeGS and SmokeSeer (Yu et al., 2026; Jain et al., 2026), with SmokeSeer additionally using thermal observations. We seek to model spatially nonuniform smoke that changes across observations while preserving a shared scene, directly from posed smoky RGB without external pretrained networks.

We propose *VeilSplat*, a physics-based Gaussian reconstruction method. Our key principle is to keep surface coverage separate from medium transmission: smoke attenuates and redistributes radiance, while the direct rendering branch keeps the scene's occlusion weights. VeilSplat concentrates capture-dependent visibility in a compact spatial and ordered-view density field. A base grid describes persistent spatial variation, and two shared basis grids describe patterns whose weights vary across the training views. Samples between each camera and Gaussian convert the field into optical thickness. A two-lobe renderer uses that thickness to attenuate direct radiance, broaden a forward-

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.

scattered contribution, and add airlight. Geometry, surface coverage and color belong to one scene throughout fitting.

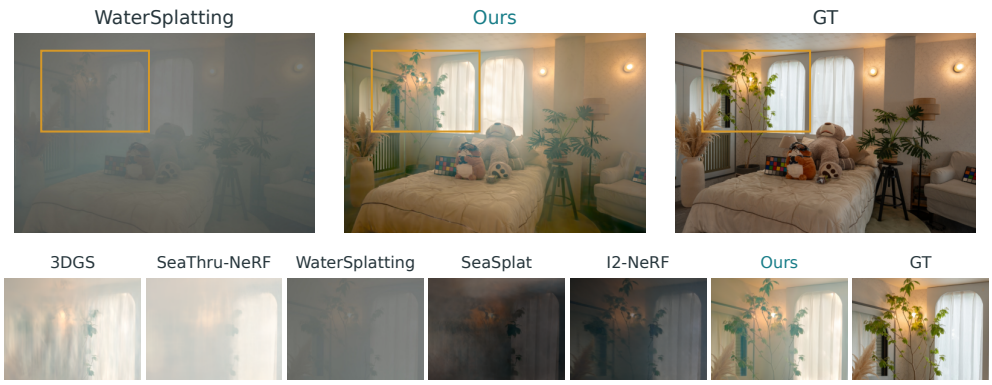


Figure 1: Clean novel views on Koharu. The marked region compares branch silhouettes and curtain folds across all five official baselines, VeilSplat and GT.

Clean inference directly renders the fitted Gaussian scene with ordinary splatting. The medium explains the training observations, while the output at any requested camera is rendered from the shared scene radiance. Figure 1 shows recovered branches and leaf silhouettes against a bright window within a complete room.

We evaluate VeilSplat on five real-world RealX3D smoke scenes. It achieves 16.30 dB mean PSNR, a 5.44 dB gain over the strongest evaluated baseline, and leads all five official baselines in mean PSNR, SSIM and LPIPS. Matched novel views show recovered window boundaries, wallpaper, upholstery and furniture across the captured rooms. Component studies connect these outcomes to the physical division of labor: coverage separation improves reconstruction, varying transmission improves PSNR and LPIPS in every scene, and forward scattering reduces observation error with the scene held fixed.

Our contributions are threefold:

- We formulate smoke-aware Gaussian rendering that separates surface coverage from medium transmission.
- We couple a compact, capture-varying spatial density field with direct, forward-scattered, and airlight rendering, enabling joint scene–medium reconstruction without external pre-trained networks.
- We demonstrate gains over five official baselines on RealX3D and isolate the roles of coverage separation, medium variation and forward scattering through matched component studies.

2 Related Work

2.1 Reconstruction in scattering media

Neural rendering uses radiance fields (Mildenhall et al., 2020; Barron et al., 2021; 2022) and hash encodings (Müller et al., 2022), with DTU and LLFF supplying multiview benchmarks (Jensen et al., 2014; Mildenhall et al., 2019). Dehazing combines atmospheric image formation (Narasimhan & Nayar, 2002; 2003; Fattal, 2008) with color priors (He et al., 2011; Berman et al., 2016) or learned transmission (Ren et al., 2016; Cai et al., 2016); D-HAZY provides paired evaluation (Ancuti et al., 2016).

Scattering-aware reconstruction separates scene radiance from intervening-medium effects. SeaThru-NeRF introduces this decomposition in volumetric rendering (Levy et al., 2023). WaterSplatting combines Gaussian scene primitives with a volumetric medium field (Li et al., 2025), while SeaSplat applies depth- and color-dependent underwater image formation (Yang et al., 2025).

I2-NeRF models physically grounded media interactions in a neural radiance field (Liu et al., 2025). DehazeGS links Gaussian depth, transmission and atmospheric light through channel priors and pseudo-depth supervision (Yu et al., 2026). VeilSplat extends this observation-model perspective to a capture-varying density field coupled to direct, broadened forward and airlight splats.

Related degraded-scene methods address blur (Ma et al., 2022; Wang et al., 2023; Cao et al., 2025), adverse illumination (Cui et al., 2024; Zhou et al., 2025; Cui et al., 2025), and restoration-guided Gaussian optimization (Lin et al., 2025; Cao et al., 2026). ReSplat learns recurrent Gaussian updates (Xu et al., 2026), while Faster-GS accelerates scene fitting (Hahlbohm et al., 2026).

2.2 Recovering a scene through changing smoke

SmokeSeer reconstructs a static scene through dynamic smoke using paired RGB and thermal observations, smoke segmentation, and separate scene and smoke representations (Jain et al., 2026). Thermal images are less affected by visible smoke. VeilSplat jointly fits the scene and medium from posed smoky RGB, using shared spatial bases and capture-dependent coefficients.

Dynamic representations track evolving geometry (Pumarola et al., 2021; Luiten et al., 2024; Yang et al., 2024), and PhysGaussian integrates material dynamics with splatting (Xie et al., 2024). Fluid reconstruction targets the evolving medium itself: FluidGS estimates fluid density and velocity from sparse videos (Xie et al., 2025), and GauSmoke combines density and motion constraints with Gaussian optimization (Zhang et al., 2026). Our target is the obstructed scene. We use Gaussian scene primitives (Kerbl et al., 2023; Kheradmand et al., 2024) and evaluate clean novel views against RealX3D's clear captures (Liu et al., 2026).

3 Method

3.1 Shared scene and observation model

Given smoky images I_v and calibrated cameras v, we fit a Gaussian scene G and medium parameters whose joint render explains each input. A Gaussian has center $\boldsymbol{\mu}_i$, covariance Σ_i, opacity α_i and color $\boldsymbol{c}_i$. Degree-zero spherical harmonics give each primitive one shared radiance value. The inputs also include calibration, sparse track correspondences and a scene-to-metric scale. The desired clean image is $J_v = \mathcal{R}_v(G)$, where $\mathcal{R}_v$ denotes ordinary Gaussian splatting.

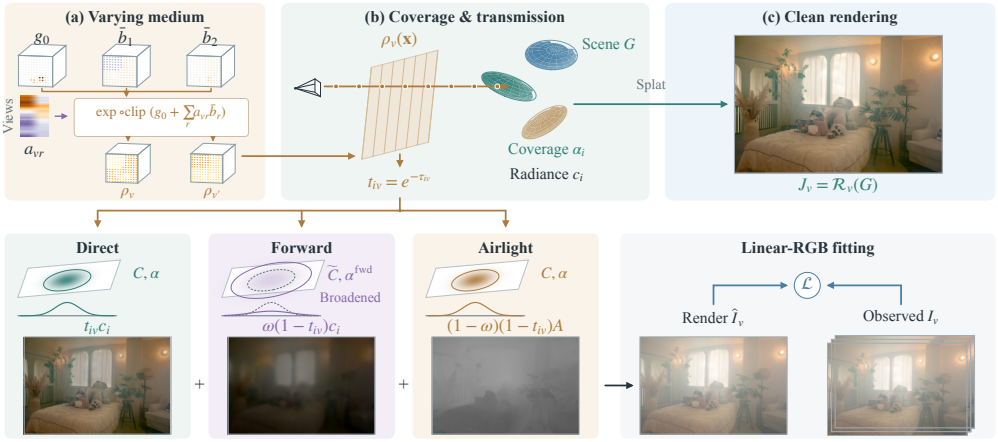


Figure 2: VeilSplat separates shared scene coverage from capture-varying smoke. Path-sampled thickness drives three observation branches; ordinary Gaussian splatting renders the clean scene.

Figure 2 separates surface coverage from transmission through smoke. Coverage determines which surfaces occlude others; transmission determines how much radiance arrives from a covered surface. Multiplying physical opacity by transmission would change both, making attenuation behave like geometric transparency. We instead retain scene opacity in the direct branch and attenuate its color.

The medium thereby changes observed brightness without giving the direct renderer an alternative surface layout.

Consider a foreground surface in front of a background object. Its coverage determines the weight passed to the background during front-to-back compositing. Attenuating the foreground color preserves that weight; attenuating its opacity increases the contribution of the background as well. The distinction therefore affects both the foreground and every surface behind it. Keeping the original occlusion weights lets the model explain a dimmed silhouette through transport while preserving the ordering of the objects that form it. Forward scattering then distributes the removed scene light through its own broadened footprint.

A locally uniform image-space model provides intuition for the two lobes:

$$I \approx tJ + (1-t)\{\omega(\kappa_\tau * J) + (1-\omega)\boldsymbol{A}\}, \qquad t = \exp(-\tau). \tag{1}$$

Optical thickness τ determines the removed fraction $1-t$. Of that fraction, ω contributes scene-dependent forward scattering through a normalized broadening kernel κ_τ, and $1-\omega$ contributes airlight $\boldsymbol{A}$. Direct and forward terms both depend on the object's radiance; airlight provides a common illumination component.

Our implementation applies this partition to primitives before compositing. This matters near boundaries: broadening a primitive changes its footprint and branch visibility, whereas convolving a completed image mixes already-resolved surfaces. The following subsections specify the density sampling and three splat passes.

3.2 Spatial density with ordered-view variation

The medium contains a base log-density grid g_0 and two spatial basis grids b_1, b_2. Each training view (capture) v selects coefficients a_{v1}, a_{v2}:

$$\ell_v(\boldsymbol{x}) = g_0(\boldsymbol{x}) + \sum_{r=1}^{2} a_{vr}\bar{b}_r(\boldsymbol{x}), \qquad \rho_v(\boldsymbol{x}) = \exp(\mathrm{clip}(\ell_v(\boldsymbol{x}), -3, 3))\,. \tag{2}$$

Each $\bar{b}_r$ is spatially centered and normalized to unit root mean square. Trilinear interpolation evaluates the grids, extending border values outside their bounds. The bounds enclose the initial Gaussian and training-camera centers with a five-percent margin on each axis.

This factorization gives different views access to the same spatial patterns. The base can represent persistent density differences across the room; a coefficient change raises or lowers a basis pattern across that room. Smoothness treats views adjacent in the sorted training-camera order as neighbors. Consequently, neighboring observations can share a medium explanation without allocating an independent volume to every image.

For primitive i and camera center $\boldsymbol{o}_v$, eight equally spaced segment midpoints give

$$\bar{\rho}_{iv} = \frac{1}{8}\sum_{k=0}^{7} \rho_v\left(\boldsymbol{o}_v + \frac{k+1/2}{8}(\boldsymbol{\mu}_i - \boldsymbol{o}_v)\right), \qquad \tau_{iv} = \sigma_t s z_{iv}\ \mathrm{clip}(\bar{\rho}_{iv}, 0.05, 20). \tag{3}$$

The positive extinction coefficient σ_t is shared across RGB channels, s converts scene units to metres, and z_{iv} is positive camera-axis depth. We exponentiate the clipped log density at each sample before averaging. Optical thickness uses this sampled density average and axial depth z_{iv}.

Density and extinction have a scale freedom: a constant log-density shift can trade against σ_t. A basis scale can similarly trade against its coefficient. After updates, we spatially center the base grid and center the coefficient matrix across views. Basis normalization fixes its sampling scale, and a Gram-matrix penalty discourages redundant bases. Spatial variation and coefficient smoothness are regularized as described below.

The field is independent of the number of Gaussians. Densification adds scene primitives that query the existing grids, rather than adding a separate smoke variable to each primitive. A view's two coefficients still induce different optical thicknesses across the image because its camera-to-primitive segments pass through different locations.

3.3 TWO-LOBE GAUSSIAN RENDERING

For transmission $t_{iv} = \exp(-\tau_{iv})$, the three primitive colors are

$$\boldsymbol{c}_{iv}^{\mathrm{dir}} = t_{iv}\boldsymbol{c}_i, \quad \boldsymbol{c}_{iv}^{\mathrm{fwd}} = \omega(1 - t_{iv})\boldsymbol{c}_i, \quad \boldsymbol{c}_{iv}^{\mathrm{air}} = (1 - \omega)(1 - t_{iv})\boldsymbol{A}. \tag{4}$$

The direct and airlight branches keep the original covariance and opacity. The forward branch broadens the covariance by adding variance to each world-space scale axis. If h is halo width in pixels at unit optical thickness, the added variance corresponds to $h^2\tau_{iv}$ in the image-plane approximation, converted using primitive depth and geometric-mean focal length. Appendix A gives this conversion explicitly.

A broadened Gaussian with unchanged peak opacity has a larger integrated footprint. We compensate its opacity using the original and broadened projected covariance matrices C_{iv} and $\widetilde{C}_{iv}$:

$$\alpha_{iv}^{\mathrm{fwd}} = \alpha_i \sqrt{\frac{\det C_{iv}}{\det \widetilde{C}_{iv}}}. \tag{5}$$

Both matrices include the rasterizer's low-pass dilation. The determinant ratio preserves the integral of an isolated projected Gaussian under the dilation. Computing it after projection accounts for viewing geometry.

Write $\mathcal{R}_v(G; \boldsymbol{q})$ for ordinary splatting after replacing primitive colors by $\boldsymbol{q}$, and let $\widetilde{G}_v$ contain the widened covariances and compensated opacities. The predicted smoky image is

$$\widehat{I}_v = \mathcal{R}_v(G; \boldsymbol{c}_v^{\mathrm{dir}}) + \mathcal{R}_v(\widetilde{G}_v; \boldsymbol{c}_v^{\mathrm{fwd}}) + \mathcal{R}_v(G; \boldsymbol{c}_v^{\mathrm{air}}). \tag{6}$$

Each pass uses a zero background, and their outputs are added in linear RGB. The primitive-level scalar weights in Equation 4 sum to one. Visibility is then resolved separately in each pass: the forward branch can spread beyond a direct footprint while the direct branch continues to use the scene's original surface coverage.

Airlight is anchored by the smoky training images. Within each image, a dark-channel operation selects candidates; the brightest candidate is retained, and the channelwise median across views defines the fixed $\boldsymbol{A}$. Extinction and halo width use positive softplus parameterizations, and the forward fraction uses a sigmoid.

3.4 JOINT FITTING AND CLEAN INFERENCE

We optimize the scene and medium in linear RGB with

$$\mathcal{L}_{\mathrm{photo}} = 0.8\,\|\widehat{I}_v - I_v\|_1 + 0.2\,[1 - \mathrm{SSIM}(\widehat{I}_v, I_v)]. \tag{7}$$

The objective also penalizes scene opacity and scale, base-grid squared magnitude and total variation, and transient coefficients and bases (the two view-weighted bases b_1, b_2 and their coefficients a_{vr}). Coefficient smoothness sums mean squared first and second differences along the ordered views. The basis penalty combines spatial total variation with the squared deviation of the normalized basis Gram matrix from the identity.

For each sampled camera, fitting retrieves its coefficients, normalizes the bases, and samples the camera-to-primitive segments. Optical thickness controls both the radiance partition and forward footprint. The three-pass residual updates scene and medium together, followed by the centering operations above. Each primitive keeps one color across views, so differences between observations must be explained by the shared medium.

A clean camera uses $J = \mathcal{R}(G)$, followed by conversion to display-referred RGB.

4 EXPERIMENTS

4.1 DATA, FITTING AND COMPARISONS

We use five RealX3D smoke scenes: Akikaze, Hinoki, Koharu, Natsume and Shirohana. Each has four clean novel viewpoints. A scene score averages its four views, and the aggregate weights the five

scenes equally. Evaluation uses the benchmark's quarter-resolution images with scene-dependent dimensions. We fit one model per scene. Initialization retriangulates sparse tracks from training observations and resamples colors from smoky training images, using the supplied calibration and track correspondences.

Every density grid has resolution 12^3, for 5,184 entries plus two coefficients per training view. Gaussian densification has a cap of one million primitives; it starts at step 600 and runs every 100 steps until step 15,000. Fitting uses 24,000 steps. Medium Adam learning rate is 0.001 and coefficient learning rate is 0.005. Initial extinction, forward fraction and halo width are 0.34, 0.22 and 29 pixels; halo width is bounded to $[0, 512]$ pixels. Airlight uses the top 0.1% dark-channel candidates with a 15×15 neighborhood. Scene opacity and scale penalties each have weight 0.01. Base-grid squared magnitude and total variation have weights 2×10^{-4} and 0.002; each transient regularizer has weight 0.001.

The image-formation reference uses scene-level extinction and fixed airlight with the same two-lobe rendering equations. Appendix B and Table 2 report its scene results.

PSNR averages the three channelwise decibel values of each display-referred RGB image in $[0, 1]$. SSIM uses data range one (Wang et al., 2004). LPIPS uses AlexNet (Zhang et al., 2018). We evaluate the released predictions of 3DGS, SeaThru-NeRF, WaterSplatting, SeaSplat and I2-NeRF at the matched clean cameras.

Component studies. We train three configurations with a common initialization, seed 0 and the schedule above. *Coupled coverage* places transmission in direct-branch opacity; *Static medium* retains separated coverage with a base density grid; *Separated + varying* uses separated coverage and two view-weighted spatial bases. Their held-out views measure reconstruction quality. A separate branch study freezes the fitted scene and refits the medium with and without forward scattering for 500 steps from the same medium initialization and camera sequence.

4.2 Comparison with released baselines

Table 1 compares VeilSplat with all five official baselines. VeilSplat reaches 16.30 dB PSNR, 0.6288 SSIM and 0.5083 LPIPS. WaterSplatting has the highest baseline mean PSNR, 10.86 dB, while 3DGS has the highest baseline mean SSIM, 0.5002. VeilSplat improves all three averages.

Table 1: Five-scene reconstruction: scene PSNR and equal-scene means.

Method	Scene PSNR (dB) ↑					Five-scene mean		
	Akikaze	Hinoki	Koharu	Natsume	Shirohana	PSNR ↑	SSIM ↑	LPIPS ↓
3DGS	11.00	7.500	11.94	<u>9.878</u>	9.069	9.878	<u>0.5002</u>	0.7626
SeaThru-NeRF	8.393	4.948	7.837	8.643	8.574	7.679	0.4653	0.8423
WaterSplatting	9.852	<u>14.13</u>	<u>14.24</u>	7.842	8.226	<u>10.86</u>	0.4441	<u>0.7479</u>
SeaSplat	<u>11.43</u>	13.67	10.00	8.629	<u>9.219</u>	10.59	0.4466	0.8480
I2-NeRF	7.396	14.06	9.689	6.162	5.704	8.602	0.2821	0.8462
Ours	**16.85**	**14.94**	**18.29**	**17.31**	**14.11**	**16.30**	**0.6288**	**0.5083**

VeilSplat leads in PSNR and SSIM on all five scenes. Appendix C gives the scene-by-method PSNR and SSIM.

Figures 1 and 3 show complementary structures: fine branches against a bright window, a window frame against a warm wall, illuminated wallpaper and striped upholstery. VeilSplat retains the object/background separation and local tonal variation in these regions. The Natsume crop adds sofa piping, buttons and a leg against parquet.

4.3 Separating coverage from transmission

Separating coverage from transmission improves mean PSNR by 1.82 dB over coupled coverage and reduces LPIPS in all five scenes (Figure 4). The intervention changes where attenuation acts: in

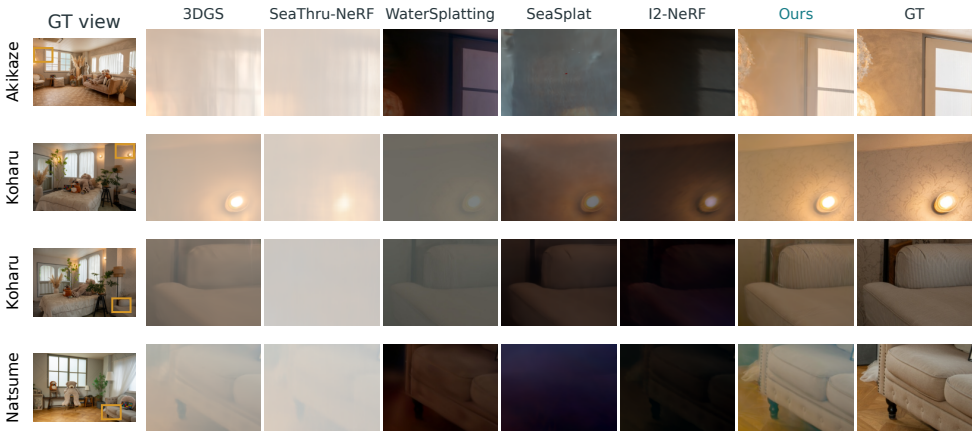


Figure 3: Window frames, wallpaper, upholstery and sofa seams. Each GT context locates the identical crop compared across all five official baselines, VeilSplat and GT.

the coupled control it alters opacity, whereas separated rendering preserves occlusion weights and attenuates radiance. The comparison thus isolates where smoke attenuation is applied.

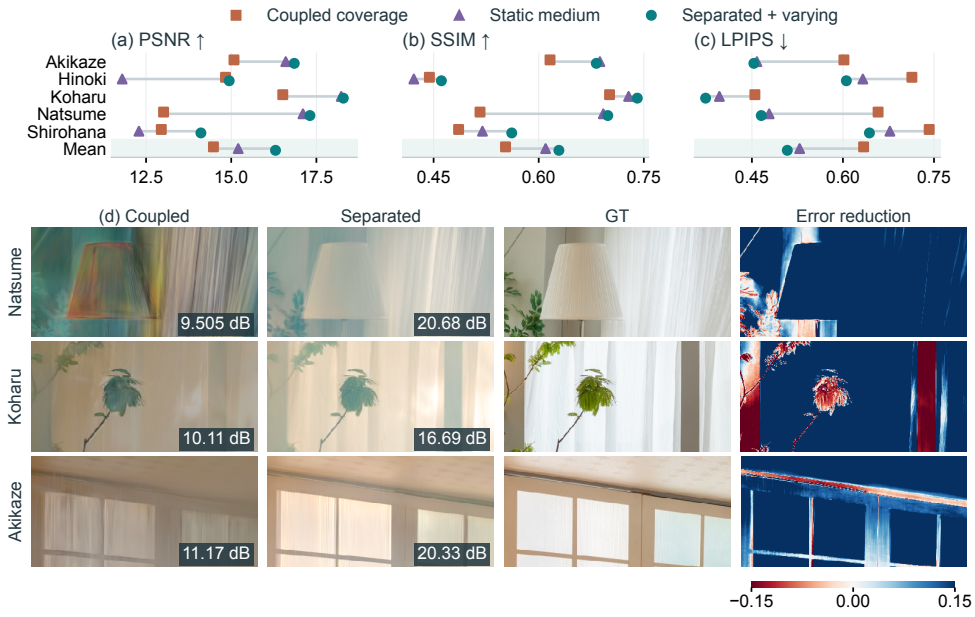


Figure 4: Coverage separation and medium variation. (a–c) Five-scene component comparisons and equal-scene means. (d) Matched boundaries: blue indicates reduced RGB error; labels report crop PSNR.

Figure 4d shows the corresponding local effect. With separated coverage, the Natsume lampshade retains a solid silhouette and the Akikaze frame retains its horizontal boundary. Shared-scale error differences locate the improvement within these structures. The same crop and display mapping are used throughout each row.

4.4 A SHARED SCENE WITH CHANGING VISIBILITY

A varying medium improves PSNR and LPIPS over a static medium in every scene and all three means (Figure 4), with a mean PSNR gain of 1.10 dB. Both configurations have a spatial density grid; the varying model adds shared spatial bases and capture coefficients.

We then hold the camera, geometry and radiance fixed and vary only the capture coefficients. Applying learned coefficients from different captures changes transmission across the same leaves, curtains and lampshade, while retaining one scene representation (Figure 5). The learned field and coefficient traces in Appendix F show how two shared spatial patterns generate these changes.

The fixed-camera sequence makes the spatial role of the field visible. Different image regions sample different paths through the same combined grids, so changing one capture coefficient induces a coordinated pattern of transmission changes across the image. The percentile bands summarize this spatial spread, while the mean traces show its evolution over capture states.

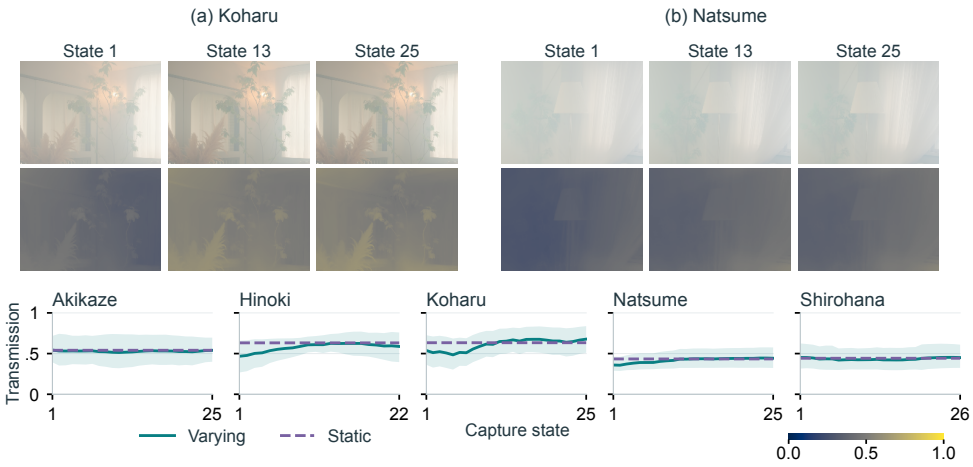


Figure 5: Changing medium at fixed cameras. Top: three RGB/transmission states for Koharu and Natsume. Bottom: all five scene traces, with spatial 10th–90th percentile bands.

4.5 OPTICAL DECOMPOSITION AND FORWARD SCATTERING

Figure 6 shows the learned image formation on shared regions in two scenes. Surface coverage stays high over the scene while transmission attenuates its radiance. Direct light retains local structure, the forward branch spreads scene-dependent light, and airlight contributes the veil. Their linear-RGB sum reproduces the smoky observation.

Frozen-scene medium refitting isolates the forward branch (Figure 7). The three-branch model achieves lower observation error at all 76 training captures of Koharu, Natsume and Shirohana. Its mean linear-RGB MSE is 25.98, 24.55 and 12.00 times lower on these scenes, respectively (Appendix G). In the Natsume window crop, the forward contribution restores the distribution of light around the frame and suppresses the corresponding residual. The medium can explain this scattered light while geometry and scene radiance remain fixed.

5 DISCUSSION

The component studies support a common design principle: assign stable appearance to the scene and changing visibility to the medium. Coverage separation preserves the direct branch's surface ordering, capture coefficients explain variations through shared spatial bases, and forward scattering accounts for redistributed scene light.

The three-branch decomposition also assigns each visible effect to a separate branch. Attenuation modulates direct radiance, broadening redistributes scene-dependent light, and airlight explains an additive veil. The fixed-scene refits show that forward scattering can be fitted through medium

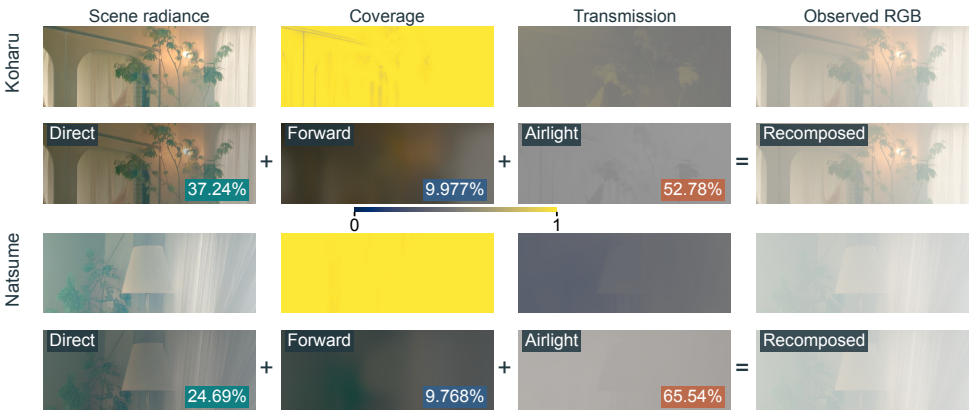


Figure 6: Optical decomposition on Koharu and Natsume. Direct radiance, forward scatter and airlight reconstruct the observation. Percentages are whole-view linear-RGB energy fractions; coverage and transmission share $[0, 1]$.

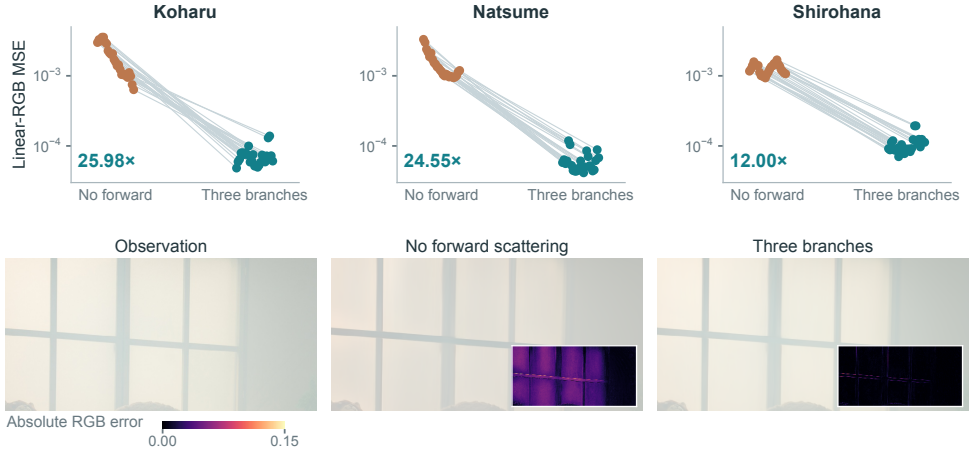


Figure 7: Forward-scattering ablation. Top: 76 paired observation errors after matched medium refits; labels give scene-mean MSE ratios. Bottom: a Natsume window crop with shared-scale RGB-error insets.

parameters while preserving the recovered objects. This division concentrates capture variation in the observation model and retains a shared appearance for novel-view rendering.

The observation model and output representation also have different costs. Training renders three branches and queries the medium; clean inference is an ordinary Gaussian render. On an RTX 4090, clean rendering takes 1.15–1.33 ms across three measured scenes at approximately 1.7 megapixels, with peak allocated memory of 0.44–0.50 GB (Appendix H, Figure 12). Any new camera pose is rendered directly from the stored scene.

6 Conclusion

VeilSplat recovers a clean shared Gaussian scene from transient smoky observations without external pretrained models by explaining RGB with a compact spatial and ordered-view medium. A path-sampled optical thickness couples that field to direct, forward-scattered and airlight splats. Across five real-world smoke scenes, VeilSplat achieves 16.30 dB mean PSNR, outperforming the strongest baseline by +5.44 dB while improving SSIM and LPIPS. Once fitting is complete, ordinary splatting renders the scene without the medium.

AI USE STATEMENT

Generative AI assisted literature search, manuscript organization and writing, interpretation of experimental results, and implementation of plotting and document preparation scripts. Experimental figures use the existing research images and measurements. Architecture diagrams are editable vector illustrations of the described method.

REFERENCES

Cosmin Ancuti, Codruta O. Ancuti, and Christophe De Vleeschouwer. D-HAZY: A dataset to evaluate quantitatively dehazing algorithms. In *2016 IEEE International Conference on Image Processing, ICIP 2016, Phoenix, AZ, USA, September 25-28, 2016*, pp. 2226–2230. IEEE, 2016. doi: 10.1109/ICIP.2016.7532754. URL https://doi.org/10.1109/ICIP.2016.7532754.

Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pp. 5835–5844. IEEE, 2021. doi: 10.1109/ICCV48922.2021.00580. URL https://doi.org/10.1109/ICCV48922.2021.00580.

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, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 5460–5469. IEEE, 2022. doi: 10.1109/CVPR52688.2022.00539. URL https://doi.org/10.1109/CVPR52688.2022.00539.

Dana Berman, Tali Treibitz, and Shai Avidan. Non-local image dehazing. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pp. 1674–1682. IEEE Computer Society, 2016. doi: 10.1109/CVPR.2016.185. URL https://doi.org/10.1109/CVPR.2016.185.

Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans. Image Process.*, 25(11):5187–5198, 2016. doi: 10.1109/TIP.2016.2598681. URL https://doi.org/10.1109/TIP.2016.2598681.

Qida Cao, Jiajun Ding, Zhenyang Liu, Zhenzhong Kuang, Yijie Shao, and Yilan Shen. VC-GS: view-consistent deblurring gaussian splatting via alternating branch optimization. *Vis. Comput.*, 41(10):7303–7317, 2025. doi: 10.1007/S00371-025-03806-Y. URL https://doi.org/10.1007/s00371-025-03806-y.

Qida Cao, Jiajun Ding, Qingyuan Tang, Tianning Zhao, Xiaoling Gu, Jianping Fan, and Zhou Yu. GC-GS: gradient control gaussian splatting with various image degradation. *Pattern Recognit.*, 171:112304, 2026. doi: 10.1016/J.PATCOG.2025.112304. URL https://doi.org/10.1016/j.patcog.2025.112304.

Ziteng Cui, Lin Gu, Xiao Sun, Xianzheng Ma, Yu Qiao, and Tatsuya Harada. Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (eds.), *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada*, pp. 1435–1444. AAAI Press, 2024. doi: 10.1609/AAAI.V38I2.27908. URL https://doi.org/10.1609/aaai.v38i2.27908.

Ziteng Cui, Xuangeng Chu, and Tatsuya Harada. Luminance-gs: Adapting 3d gaussian splatting to challenging lighting conditions with view-adaptive curve adjustment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 26472–26482. Computer Vision Foundation / IEEE, 2025. doi: 10.1109/CVPR52734.2025.02465. URL https://openaccess.thecvf.com/content/CVPR2025/html/Cui_Luminance-GS_Adapting_3D_Gaussian_Splatting_to_Challenging_Lighting_Conditions_with_CVPR_2025_paper.html.

Raanan Fattal. Single image dehazing. *ACM Trans. Graph.*, 27(3):72, 2008. doi: 10.1145/1360612.1360671. URL https://doi.org/10.1145/1360612.1360671.

Florian Hahlbohm, Linus Franke, Martin Eisemann, and Marcus A. Magnor. Faster-gs: Analyzing and improving gaussian splatting optimization. *CoRR*, abs/2602.09999, 2026. doi: 10.48550/ARXIV.2602.09999. URL https://doi.org/10.48550/arXiv.2602.09999.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(12):2341–2353, 2011. doi: 10.1109/TPAMI.2010.168. URL https://doi.org/10.1109/TPAMI.2010.168.

Neham Jain, Andrew Jong, Sebastian A. Scherer, and Ioannis Gkioulekas. Smokeseer: 3d gaussian splatting for smoke removal and scene reconstruction. In *International Conference on 3D Visio, 3DV 2026, Vancouver, BC, Canada, March 20-23, 2026*, pp. 510–519. IEEE, 2026. doi: 10.1109/3DV69130.2026.00055. URL https://doi.org/10.1109/3DV69130.2026.00055.

Rasmus Ramsbøl Jensen, Anders Lindbjerg Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pp. 406–413. IEEE Computer Society, 2014. doi: 10.1109/CVPR.2014.59. URL https://doi.org/10.1109/CVPR.2014.59.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139:1–139:14, 2023. doi: 10.1145/3592433. URL https://doi.org/10.1145/3592433.

Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3d gaussian splatting as markov chain monte carlo. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/93be245fce00a9bb2333c17ceae4b732-Abstract-Conference.html.

Deborah Levy, Amit Peleg, Naama Pearl, Dan Rosenbaum, Derya Akkaynak, Simon Korman, and Tali Treibitz. Seathru-nerf: Neural radiance fields in scattering media. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pp. 56–65. IEEE, 2023. doi: 10.1109/CVPR52729.2023.00014. URL https://doi.org/10.1109/CVPR52729.2023.00014.

Huapeng Li, Wenxuan Song, Tianao Xu, Alexandre Elsig, and Jonas Kulhanek. Watersplatting: Fast underwater 3d scene reconstruction using gaussian splatting. In *International Conference on 3D Vision, 3DV 2025, Singapore, March 25-28, 2025*, pp. 969–978. IEEE, 2025. doi: 10.1109/3DV66043.2025.00094. URL https://doi.org/10.1109/3DV66043.2025.00094.

Xin Lin, Shi Luo, Xiaojun Shan, Xiaoyu Zhou, Chao Ren, Lu Qi, Ming-Hsuan Yang, and Nuno Vasconcelos. HQGS: high-quality novel view synthesis with gaussian splatting in degraded scenes. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL https://openreview.net/forum?id=25Zlvl7JxW.

Shuhong Liu, Lin Gu, Ziteng Cui, Xuangeng Chu, and Tatsuya Harada. I2-nerf: Learning neural radiance fields under physically-grounded media interactions. In Danielle Belgrave, Cheng Zhang, Laura N. Montoya, Hsuan-Tien Lin, Razvan Pascanu, Piotr Koniusz, Marzyeh Ghassemi, Nancy Chen, Iván Vladimir Meza Ruíz, and Arturo Loaiza-Bonilla (eds.), *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2025, NeurIPS 2025, San Diego, CA, USA, December 2-7, 2025 / Mexico City, Mexico, November 30 - December 5, 2025*, 2025. URL http://papers.nips.cc/paper_files/paper/2025/hash/77fa0e7d45c6687f1958de0b31e9fc05-Abstract-Conference.html.

Shuhong Liu, Chenyu Bao, Ziteng Cui, Yun Liu, Xuangeng Chu, Lin Gu, Marcos V. Conde, Ryo Umagami, Tomohiro Hashimoto, Zijian Hu, Tianhan Xu, Yuan Gan, Yusuke Kurose, and Tatsuya Harada. Realx3d: A physically-degraded 3d benchmark for multi-view visual restoration and reconstruction. *Int. J. Comput. Vis.*, 134(8):382, 2026. doi: 10.1007/S11263-026-02963-4. URL https://doi.org/10.1007/s11263-026-02963-4.

Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *International Conference on 3D Vision, 3DV 2024, Davos, Switzerland, March 18-21, 2024*, pp. 800–809. IEEE, 2024. doi: 10.1109/3DV62453.2024.00044. URL https://doi.org/10.1109/3DV62453.2024.00044.

Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V. Sander. Deblur-nerf: Neural radiance fields from blurry images. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 12851–12860. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01252. URL https://doi.org/10.1109/CVPR52688.2022.01252.

Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: practical view synthesis with prescriptive sampling guidelines. *ACM Trans. Graph.*, 38(4):29:1–29:14, 2019. doi: 10.1145/3306346.3322980. URL https://doi.org/10.1145/3306346.3322980.

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 Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (eds.), *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I*, volume 12346 of *Lecture Notes in Computer Science*, pp. 405–421. Springer, 2020. doi: 10.1007/978-3-030-58452-8_24. URL https://doi.org/10.1007/978-3-030-58452-8_24.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, 2022. doi: 10.1145/3528223.3530127. URL https://doi.org/10.1145/3528223.3530127.

Srinivasa G. Narasimhan and Shree K. Nayar. Vision and the atmosphere. *Int. J. Comput. Vis.*, 48(3): 233–254, 2002. doi: 10.1023/A:1016328200723. URL https://doi.org/10.1023/A:1016328200723.

Srinivasa G. Narasimhan and Shree K. Nayar. Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(6):713–724, 2003. doi: 10.1109/TPAMI.2003.1201821. URL https://doi.org/10.1109/TPAMI.2003.1201821.

Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pp. 10318–10327. Computer Vision Foundation / IEEE, 2021. doi: 10.1109/CVPR46437.2021.01018. URL https://openaccess.thecvf.com/content/CVPR2021/html/Pumarola_D-NeRF_Neural_Radiance_Fields_for_Dynamic_Scenes_CVPR_2021_paper.html.

Wenqi Ren, Si Liu, Hua Zhang, Jin-shan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (eds.), *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*, volume 9906 of *Lecture Notes in Computer Science*, pp. 154–169. Springer, 2016. doi: 10.1007/978-3-319-46475-6_10. URL https://doi.org/10.1007/978-3-319-46475-6_10.

Peng Wang, Lingzhe Zhao, Ruijie Ma, and Peidong Liu. Bad-nerf: Bundle adjusted deblur neural radiance fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pp. 4170–4179. IEEE, 2023. doi: 10.1109/CVPR52729.2023.00406. URL https://doi.org/10.1109/CVPR52729.2023.00406.

Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861. URL https://doi.org/10.1109/TIP.2003.819861.

Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 4389–4398. IEEE, 2024. doi: 10.1109/CVPR52733.2024.00420. URL https://doi.org/10.1109/CVPR52733.2024.00420.

Youchen Xie, Chen Li, Sheng Qiu, Zhi-Jun Wang, Chenhui Li, Yibo Zhao, Zan Gao, and Changbo Wang. Fluidgs: Physics informed gaussian splatting for dynamic fluid reconstruction from sparse views. In Cathal Gurrin, Klaus Schoeffmann, Min Zhang, Luca Rossetto, Stevan Rudinac, Duc-Tien Dang-Nguyen, Wen-Huang Cheng, Phoebe Chen, and Jenny Benois-Pineau (eds.), *Proceedings of the 33rd ACM International Conference on Multimedia, MM 2025, Dublin, Ireland, October 27-31, 2025*, pp. 8438–8447. ACM, 2025. doi: 10.1145/3746027.3755500. URL https://doi.org/10.1145/3746027.3755500.

Haofei Xu, Daniel Barath, Andreas Geiger, and Marc Pollefeys. Resplat: Learning recurrent gaussian splatting. In Paolo Favaro, Zuzana Kukelova, Atsuto Maki, Anna Rohrbach, Konrad Schindler, and Federico Tombari (eds.), *Computer Vision - ECCV 2026 - 19th European Conference, Malmö, Sweden, September 8-12, 2026, Proceedings, Part XXXIX*, volume 17039 of *Lecture Notes in Computer Science*, pp. 364–382. Springer, 2026. doi: 10.1007/978-3-032-37435-6_20. URL https://doi.org/10.1007/978-3-032-37435-6_20.

Daniel Yang, John J. Leonard, and Yogesh A. Girdhar. Seasplat: Representing underwater scenes with 3d gaussian splatting and a physically grounded image formation model. In *IEEE International Conference on Robotics and Automation, ICRA 2025, Atlanta, GA, USA, May 19-23, 2025*, pp. 7632–7638. IEEE, 2025. doi: 10.1109/ICRA55743.2025.11128502. URL https://doi.org/10.1109/ICRA55743.2025.11128502.

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, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 20331–20341. IEEE, 2024. doi: 10.1109/CVPR52733.2024.01922. URL https://doi.org/10.1109/CVPR52733.2024.01922.

Jinze Yu, Yiqun Wang, Aiheng Jiang, Zhengda Lu, Jianwei Guo, Yong Li, Hongxing Qin, and Xiaopeng Zhang. Dehazegs: Seeing through fog with 3d gaussian splatting. In Sven Koenig, Chad Jenkins, and Matthew E. Taylor (eds.), *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, pp. 12153–12161. AAAI Press, 2026. doi: 10.1609/AAAI.V40I14.38205. URL https://doi.org/10.1609/aaai.v40i14.38205.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pp. 586–595. Computer Vision Foundation / IEEE Computer Society, 2018. doi: 10.1109/CVPR.2018.00068. URL http://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.

Wenran Zhang, Yuxiang Cai, Letian Huang, Dongwei Ye, Jie Guo, and Bo Ren. Gausmoke: Hybrid physics-optical gaussian splatting for sparse smoke reconstruction. In Christopher P. Redmann, Justin Solomon, and Mirela Ben-Chen (eds.), *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference, SIGGRAPH Conference Papers 2026, Los Angeles, CA, USA, July 19-23, 2026*, pp. 51:1–51:11. ACM, 2026. doi: 10.1145/3799902.3811148. URL https://doi.org/10.1145/3799902.3811148.

Han Zhou, Wei Dong, and Jun Chen. LITA-GS: illumination-agnostic novel view synthesis via reference-free 3d gaussian splatting and physical priors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 21580–21589. Computer Vision Foundation / IEEE, 2025. doi: 10.1109/CVPR52734.2025.02010. URL https://openaccess.thecvf.com/content/CVPR2025/html/Zhou_LITA-GS_Illumination-Agnostic_Novel_View_Synthesis_via_Reference-Free_3D_Gaussian_Splatting_CVPR_2025_paper.html.

A Implementation Details

For log scale q_{ij} on world-space axis j, the forward branch uses

$$\widetilde{q}_{iv,j} = \tfrac{1}{2}\log\left(\exp(2q_{ij}) + h^2\tau_{iv}(z_{iv}/f_v)^2 + \epsilon\right), \quad f_v = \sqrt{f_{xv}f_{yv}}. \tag{8}$$

The isotropic world-space variance need not project to an isotropic image kernel away from the optical axis. Equation 5 therefore uses the projected covariances from the camera rotation and projective Jacobian. Both include the rasterizer's 0.3-pixel dilation term. The compensation factor is clamped to $[0, 1]$, and activated opacity is clamped away from zero and one before conversion back to logits.

B Complete Scene Results

The image-formation reference uses scene-level extinction and fixed airlight. VeilSplat improves PSNR, SSIM and LPIPS in each scene and in the equal-scene mean (Table 2). Each scene averages four clean views. Figures 8 and 9 compare the two at matched crops and at all twenty evaluation cameras.

Table 2: Image-formation reference and VeilSplat: scene results and equal-scene means.

Scene	Variant	PSNR (dB) ↑	SSIM ↑	LPIPS ↓
Akikaze	Image formation	15.21	0.6712	0.4914
	Ours	**16.85**	**0.6822**	**0.4532**
Hinoki	Image formation	9.608	0.3672	0.6593
	Ours	**14.94**	**0.4611**	**0.6056**
Koharu	Image formation	16.84	0.7180	0.4169
	Ours	**18.29**	**0.7407**	**0.3737**
Natsume	Image formation	14.25	0.6792	0.5189
	Ours	**17.31**	**0.6986**	**0.4654**
Shirohana	Image formation	11.48	0.5084	0.6853
	Ours	**14.11**	**0.5615**	**0.6436**
Mean	Image formation	13.48	0.5888	0.5544
	Ours	**16.30**	**0.6288**	**0.5083**

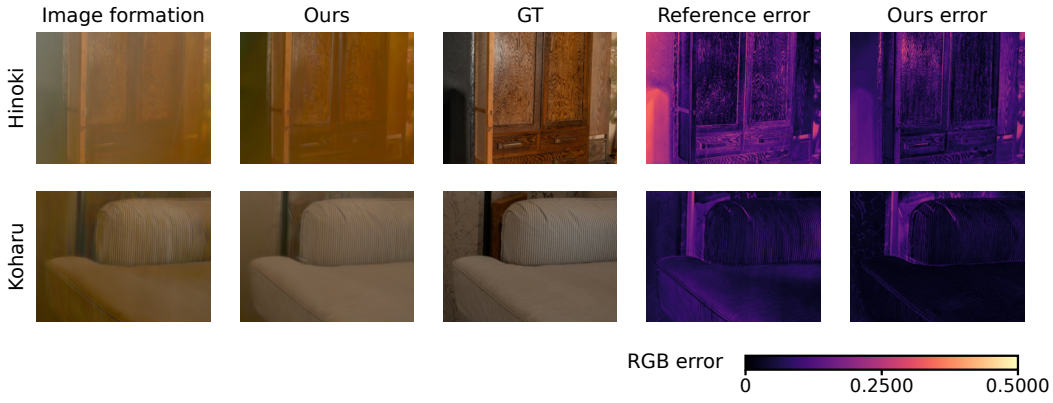


Figure 8: Matched reconstruction details and absolute RGB errors. Image-formation reference, VeilSplat and GT use identical crops; both error columns share one scale.

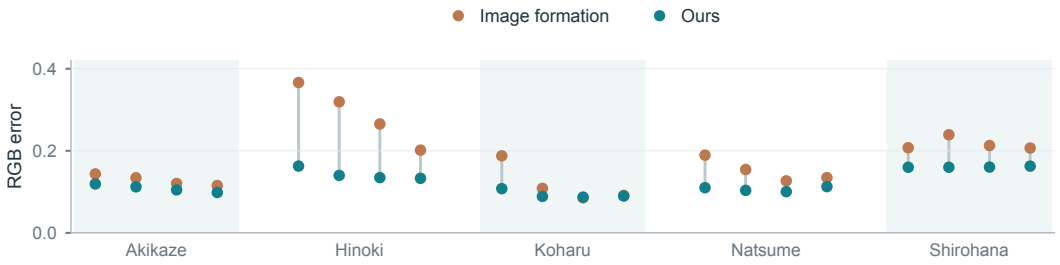


Figure 9: RGB errors at all twenty evaluation cameras. Each segment connects the image-formation reference and VeilSplat at one camera.

C Scene-level External Comparison

Table 3 gives the scene values underlying Table 1. Each scene averages four matched clean cameras; the mean weights scenes equally.

Table 3: External PSNR and SSIM, grouped by metric. Ranks compare methods within one scene and metric.

Method	Akikaze	Hinoki	Koharu	Natsume	Shirohana	Mean
			PSNR (dB) ↑			
3DGS	11.00	7.500	11.94	9.878	9.069	9.878
SeaThru-NeRF	8.393	4.948	7.837	8.643	8.574	7.679
WaterSplatting	9.852	14.13	14.24	7.842	8.226	10.86
SeaSplat	11.43	13.67	10.00	8.629	9.219	10.59
I2-NeRF	7.396	14.06	9.689	6.162	5.704	8.602
Ours	**16.85**	**14.94**	**18.29**	**17.31**	**14.11**	**16.30**
			SSIM ↑			
3DGS	0.5489	0.3221	0.6068	0.6029	0.4204	0.5002
SeaThru-NeRF	0.5135	0.2658	0.5392	0.5940	0.4140	0.4653
WaterSplatting	0.4330	0.4393	0.6368	0.3719	0.3392	0.4441
SeaSplat	0.4985	0.4277	0.4958	0.4620	0.3492	0.4466
I2-NeRF	0.2521	0.4095	0.3951	0.2155	0.1382	0.2821
Ours	**0.6822**	**0.4611**	**0.7407**	**0.6986**	**0.5615**	**0.6288**

D Recovered Boundaries and Surface Detail

Figure 10 compares the Hinoki television and Shirohana brickwork with WaterSplatting. The enlarged regions reveal the television's bezel, dial and grille, and the brick wall's mortar joints. VeilSplat resolves these boundaries and surface patterns in the reconstructed scene.

E Matched Component Results and Medium Variation

The coverage and medium study trains all three configurations with seed 0. The coupled configuration uses $t\alpha$ in the direct branch with unattenuated scene color; the separated configuration uses α with color $t\boldsymbol{c}$. The static medium sets the transient rank to zero and retains the base grid. The separated + varying configuration is the full VeilSplat model of Table 1. The scene representation, initialization, optimization schedule and remaining losses are shared. Table 4 lists the complete results underlying Figure 4.



Figure 10: Recovered television trim in Hinoki and brickwork in Shirohana. GT context views locate the shared detail crops.

Table 4: Matched component study. PSNR is in dB; SSIM and LPIPS follow the main evaluation. Each scene averages four held-out views.

	Coupled coverage			Static medium			Separated + varying		
Scene	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Akikaze	15.08	0.6161	0.6016	16.60	0.6873	0.4582	16.85	0.6822	0.4532
Hinoki	14.83	0.4440	0.7136	11.80	0.4218	0.6328	14.94	0.4611	0.6056
Koharu	16.51	0.7011	0.4549	18.23	0.7283	0.3964	18.29	0.7407	0.3737
Natsume	13.01	0.5162	0.6579	17.11	0.6922	0.4784	17.31	0.6986	0.4654
Shirohana	12.95	0.4856	0.7421	12.29	0.5197	0.6774	14.11	0.5615	0.6436
Mean	14.48	0.5526	0.6340	15.20	0.6099	0.5286	16.30	0.6288	0.5083

For the fixed-camera visualization in Figure 5, we use the middle training camera and learned coefficients from captures 1, 13 and 25. Geometry and radiance remain fixed. The transmission curves average the full rendered view; the image panels use a common detail crop within each scene and a shared $[0, 1]$ transmission scale.

F Learned Representation and Optical Branches

Figure 11 visualizes a common depth slice of Koharu's learned base and basis grids. The two coefficient traces combine these patterns into a different density field at each capture. The fixed-camera transmission distribution changes with those coefficients, linking the compact representation to its image-space response.

The direct, forward-scattered and airlight branches in Figure 6 sum to the observation render in linear RGB, with maximum absolute numerical discrepancy 1.192×10^{-7}. RGB panels use standard sRGB encoding. Their distinct spatial patterns show the scene-dependent broadening contributed by forward scattering alongside attenuated direct light and airlight.

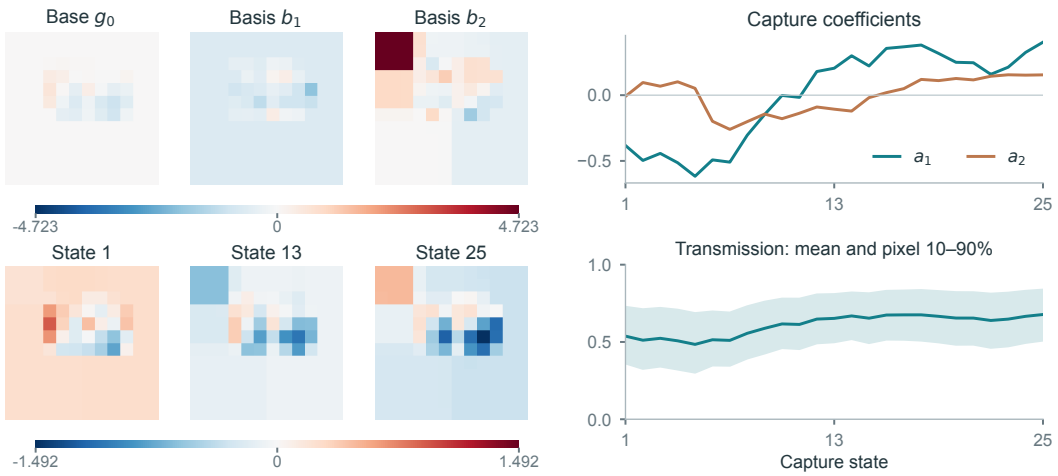


Figure 11: Koharu medium representation: base density, spatial bases, capture coefficients and three combined states. Fixed-camera transmission bands span the spatial 10th–90th percentiles.

G Matched Medium Refitting

The branch study fixes Gaussian geometry and radiance and optimizes medium parameters and capture coefficients with Adam at learning rate 0.001. The two arms use the same $0.8L_1 + 0.2(1 - \mathrm{SSIM})$ objective on smoky training views. Koharu and Natsume each contribute 25 captures, and Shirohana contributes 26. Figure 7 pairs their post-refit errors.

After refitting, the complete model achieves mean linear-RGB MSE of 7.211×10^{-5}, 6.000×10^{-5} and 1.055×10^{-4} on Koharu, Natsume and Shirohana, respectively. The corresponding models without forward scattering reach 1.874×10^{-3}, 1.473×10^{-3} and 1.266×10^{-3}. The complete model has lower error at all 76 captures, with scene-mean error ratios of 25.98, 24.55 and 12.00.

H Rendering Cost

We measure rendering on one RTX 4090 using CUDA events. Each scene uses its first, middle and last training cameras, five warm-up calls and 100 timed calls per camera. Image dimensions are 1599×1059, 1600×1065 and 1600×1054 for Koharu, Natsume and Shirohana. We report the median of the three camera medians and the maximum peak allocated GPU memory across those cameras; reference images reside on the CPU. Clean renders take 1.333, 1.165 and 1.146 ms, respectively. Their peak allocated memory is 0.4969, 0.4438 and 0.4388 GB.

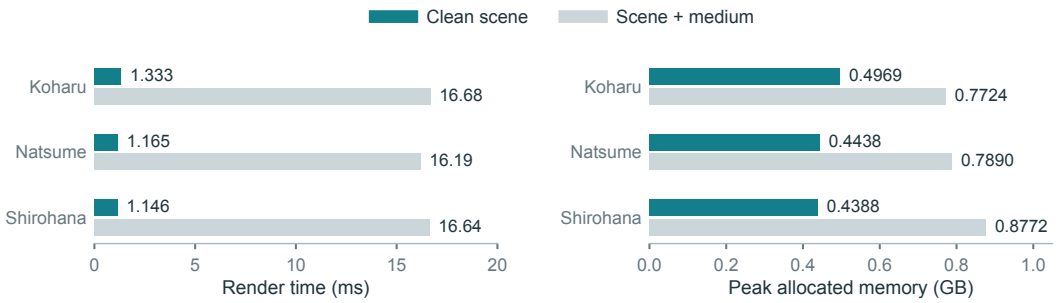


Figure 12: Rendering latency and peak allocation. Clean views use standard Gaussian splatting; observation rendering also evaluates the medium and three branches.