ONE FRAMEWORK, DIVERSE INFORMATION DEFICIENCIES: TOWARDS UNIFIED 3D GAUSSIAN RECONSTRUCTION UNDER IMPERFECT OBSERVATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Real-world 3D reconstruction is often limited by diverse imperfect observations, ranging from incomplete scene coverage to degraded image quality. Despite their different manifestations, we argue that they share a common underlying problem: insufficient reliable scene information for faithful reconstruction. We unify them under the notion of **information deficiency** and introduce **UniFill3D**, a unified framework for 3D Gaussian reconstruction under imperfect observations. UniFill3D couples scene-grounded reconstruction, which preserves scene-specific geometry and multi-view consistency, with external prior recovery, which recovers missing, degraded, or unreliable visual information. Their target-aligned predictions are reconciled through hierarchical scene–prior fusion, combining dual-latent single-step diffusion with multi-scale prior-guided decoding. The resulting restored views are then fed back to refine the 3D Gaussian representation. We validate UniFill3D on four representative settings spanning incomplete and degraded observations: sparse-view reconstruction, 3D super-resolution, 3D deblurring, and 3D low-light enhancement. UniFill3D achieves state-of-the-art performance across all four tasks, with ablations confirming the complementary value of scene grounding and external priors. Together, these results support viewing seemingly disparate 3D reconstruction problems as different manifestations of a shared information-deficiency challenge that can be addressed through a common reconstruction paradigm.

1 INTRODUCTION

Real-world 3D reconstruction rarely operates under ideal observations. Scene coverage may be incomplete (Zhu et al., 2024; Zhou et al., 2025b), while available images may be low-resolution (Feng et al., 2026), blurred Zhao et al. (2024); Lee et al. (2024a), poorly illuminated Cui et al. (2025), or otherwise degraded. These conditions are usually treated as separate reconstruction problems, leading to specialized methods for each task. We argue that despite their different manifestations, they share a common bottleneck: insufficient reliable scene information for faithful reconstruction.

We refer to this common challenge as *information deficiency*: useful scene information is either absent from the observation or corrupted within observations. This perspective unifies seemingly distinct tasks bound by the same fundamental uncertainty. It also changes the question we ask. Can one reconstruction mechanism address diverse forms of information deficiency? Existing methods usually couple information recovery to the specific deficiency. For example, sparse-view methods compensate for missing viewpoints (Xu et al., 2025; Li et al., 2024), super-resolution methods (Wan et al., 2025; Asthana et al., 2026) recover fine details, and deblurring methods model blur (Peng et al., 2024; Kong et al., 2026). Even when learned generative or restoration priors are introduced, the recovery mechanism and its interaction with the 3D representation remain largely task-specific. This fragmentation makes it difficult to reuse an existing reconstruction mechanism across heterogeneous observation failures.

Our core insight is that these task-specific differences can be isolated before information integration. For a target camera, imperfect observations can provide two complementary estimates of the same view: a *scene-grounded reconstruction*, anchored to the current 3D representation, and an *external*

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.



Figure 1: **UniFill3D for 3D Gaussian reconstruction under diverse information deficiencies.** Top: representative novel-view comparisons for sparse-view reconstruction, 3D super-resolution, 3D deblurring, and 3D low-light enhancement, with enlarged regions highlighting local details. Bottom: these seemingly different tasks correspond to different forms of information deficiency, missing observations, limited resolution, blur, and low illumination, which UniFill3D addresses within an unified reconstruction framework.

prior prediction, produced by a learned generative or restoration model. The former preserves scene-specific geometry and multi-view consistency but is limited by the available observations; the latter can recover missing or degraded visual information but may deviate from the true scene. Once both are expressed on the same target camera and image plane, the downstream problem becomes common: how to reconcile scene-grounded evidence with external priors, and how to consolidate the restored views back into a coherent 3D representation.

Based on this abstraction, we propose **UniFill3D**, a unified framework for 3D Gaussian reconstruction under diverse information deficiencies. UniFill3D separates deficiency-specific information recovery from shared information reconciliation. A Reconstruction Anchor and a Prior Prediction are first produced for the same target view and then reconciled by Hierarchical Scene–Prior Fusion, which combines dual-latent interaction within a single-step diffusion process with multi-scale prior-guided decoding. The restored views are subsequently fed back through Restoration-Guided 3D Refinement. Different tasks may use different reconstruction and prior models, while sharing the same information-deficiency formulation, target-aligned fusion interface, and restoration-to-3D refinement mechanism. We validate UniFill3D on four heterogeneous settings: sparse-view reconstruction, 3D super-resolution, 3D deblurring, and 3D low-light enhancement. UniFill3D achieves state-of-the-art performance across all four tasks, while ablations confirm the complementary value of scene-grounded reconstruction and external priors.

Our contributions are threefold:

- We formulate incomplete and degraded observations under *information-deficiency* perspective, connecting several traditionally separate 3D reconstruction problems.
- We propose UniFill3D, which maps heterogeneous deficiencies to a common pair of target-aligned scene-grounded and prior-based predictions, and reconciles them through a shared hierarchical fusion and 3D refinement mechanism.
- We demonstrate the generality of UniFill3D across four heterogeneous reconstruction settings, achieving state-of-the-art performance throughout.

2 Related Work

2.1 3D Reconstruction under Imperfect Observations

3D reconstruction under imperfect observations has largely been studied through separate task-specific formulations. For sparse inputs, NeRF- and Gaussian-based methods use unseen-view regularization, semantic or entropy constraints, monocular depth priors, and co-regularization (Niemeyer et al., 2021; Chung et al., 2024; Li et al., 2024; Zhu et al., 2024; Zhang et al., 2024). Feed-forward approaches such as pixelSplat, MVSplat, and DepthSplat instead learn to infer 3D Gaussians directly from sparse observations (Charatan et al., 2024; Chen et al., 2024; Xu et al., 2025). For resolution-limited inputs, representative methods include NeRF-SR, Mip-Splatting, and SRGS (Wang et al., 2022; Yu et al., 2023; Feng et al., 2026). Deblurring and low-light reconstruction are similarly addressed by dedicated methods such as Deblurring-3DGS, BAD-Gaussians, and BAGS, Luminance-GS, LITA-GS, and LL-Gaussian. Despite strong progress within each setting, these problems are predominantly formulated and solved in isolation, with assumptions and optimization strategies tailored to a particular observation deficiency. We instead view missing and degraded observations as instances of a common information-deficiency problem.

2.2 Generative and Restoration Priors

Generative and restoration models provide a complementary source of information by recovering scene content that is missing or corrupted in the observations. For unseen-view prediction, 3DiM and GeNVS introduced pose- and geometry-conditioned diffusion (Watson et al., 2022; Chan et al., 2023), followed by models such as ZeroNVS, CAT3D, and SEVA that support increasingly flexible novel-view generation from sparse observations (Sargent et al., 2024; Gao et al., 2024; Zhou et al., 2025b). For image super-resolution, StableSR, PASD, SeeSR, DiffBIR, and SUPIR exploit large-scale generative priors to recover high-frequency visual content (Wang et al., 2024; Yang et al., 2024; Wu et al., 2024b; Lin et al., 2024; Yu et al., 2024a), while SinSR and OSEDiff further enable efficient single-step restoration (Wang et al., 2023; Wu et al., 2024a). Image deblurring has likewise benefited from strong learned restoration models, including Restormer, DiffIR, and diffusion-based DeblurDiff (Zamir et al., 2022; Xia et al., 2023; Kong et al., 2026). For low-light enhancement, methods such as Retinexformer and LightenDiffusion recover visibility (Cai et al., 2023; Jiang et al., 2024), contrast, and appearance under severe illumination degradation. Although their architectures and conditioning signals differ, these models share an important capability: they can infer visual information that is weak or unavailable in the observations themselves. However, such recovery is primarily performed in image or view space and does not inherently guarantee consistency with the geometry and appearance of a particular reconstructed scene.

2.3 Integrating Image Priors with 3D Reconstruction

Learned image priors have been integrated into 3D reconstruction through optimization signals, intermediate representations, or generated observations. DiffusioNeRF and Nerfbusters introduce diffusion priors through reconstruction objectives, while GaussianSR transfers a 2D super-resolution prior into 3D Gaussians (Wynn & Turmukhambetov, 2023; Warburg et al., 2023; Yu et al., 2024b). More tightly coupled approaches include SparseFusion, ReconFusion, and ReconX, which use view-conditioned diffusion or generated views to improve sparse-view reconstruction (Zhou & Tulsiani, 2023; Wu et al., 2023; Liu et al., 2025), as well as Super-NeRF and DiSR-NeRF for 3D super-resolution (Han et al., 2023; Lee et al., 2024b).

More closely related to our work are methods that restore or generate observations and feed them back into reconstruction. Deceptive-NeRF/3DGS augments sparse inputs with diffusion-generated pseudo-views (Liu et al., 2024b); 3DGS-Enhancer, GenFusion, and DiFiX3D+ refine scenes using restored or generated observations (Liu et al., 2024a; Wu et al., 2025b;a); and GSFixer and Geo-Query further exploit reference information for sparse-view reconstruction (Yin et al., 2025; Cao et al., 2026). A parallel line in 3D super-resolution uses enhanced observations to refine the scene representation (Shen et al., 2024; Chen et al., 2025; Wan et al., 2025; Feng et al., 2025; Asthana et al., 2026). These works demonstrate the value of coupling strong image priors with 3D reconstruction, but are largely designed for a specific reconstruction problem or degradation. UniFill3D instead adopts a common information-deficiency formulation and a shared mechanism of scene-grounded

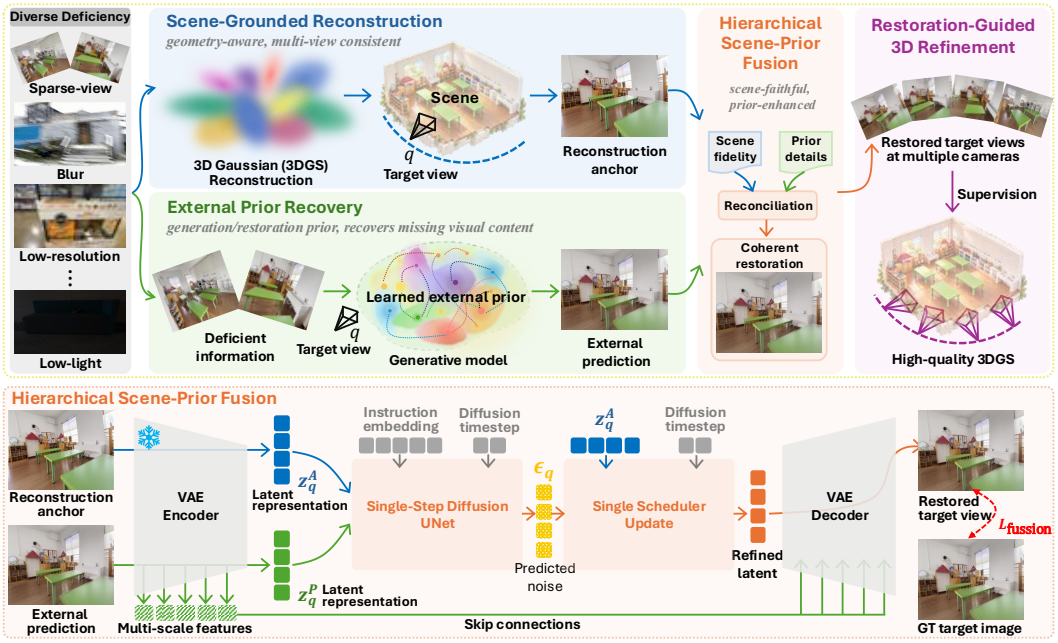


Figure 2: **Overview of UniFill3D.** Top: UniFill3D addresses diverse information-deficient 3DGS reconstructions through a common pipeline of Scene-Grounded Reconstruction, External Prior Recovery, Hierarchical Scene–Prior Fusion, and Restoration-Guided 3D Refinement. Bottom: the fusion module reconciles the target-aligned Reconstruction Anchor and Prior Prediction through joint dual-latent interaction, a single-step diffusion update, and multi-scale prior-guided decoding.

reconstruction, external prior recovery, hierarchical scene–prior fusion, and restoration-guided 3D refinement across heterogeneous observation conditions.

3 METHOD

3.1 PROBLEM FORMULATION AND FRAMEWORK OVERVIEW

We consider 3D reconstruction from imperfect observations$\mathcal{O} = \{(I_i, \pi_i)\}_{i=1}^{N}$, where I_i denotes an observed image and π_i its camera. We use *information deficiency* to describe settings in which $\mathcal{O}$ provides insufficient reliable scene information, either because relevant observations are missing or because the available images are degraded.

UniFill3D addresses different forms of information deficiency through a common scene–prior formulation (Fig. 2). Given a target camera q, Scene-Grounded Reconstruction produces a Reconstruction Anchor A_q, while External Prior Recovery produces a complementary Prior Prediction P_q:

$$A_q = \mathcal{R}(\mathcal{O}, q), \qquad P_q = \mathcal{P}(\mathcal{O}, q). \tag{1}$$

Here, $\mathcal{R}$ is grounded in the current scene representation, whereas $\mathcal{P}$ draws on a learned generative or restoration prior. Their target-aligned predictions are reconciled by Hierarchical Scene–Prior Fusion,

$$\widehat{I}_q = \mathcal{F}_\phi(A_q, P_q), \tag{2}$$

yielding a restored target view $\widehat{I}_q$. The restored view–camera pairs are then used as improved supervision to refine the underlying Gaussian representation. This defines a common pipeline of scene grounding, prior recovery, scene–prior reconciliation, and 3D refinement across heterogeneous information deficiencies.

3.2 TARGET-ALIGNED SCENE AND PRIOR PREDICTIONS

The two prediction routes provide complementary information about the same target view. The Reconstruction Anchor A_q is rendered from a 3D representation reconstructed from the available observations. It therefore preserves scene-specific geometry and multi-view consistency, but remains limited by information missing or corrupted in $\mathcal{O}$. In contrast, the Prior Prediction P_q uses a learned generative or restoration model to recover information unavailable from the observations alone, but may introduce structures or details inconsistent with the actual scene.

A key design is to expose these heterogeneous sources through a common target-aligned interface. Both A_q and P_q estimate the view at camera q on the same image plane, so corresponding locations refer to the same target rays. Target alignment does not require the two predictions to agree: they may differ in structure, texture, or reconstruction artifacts. Instead, it isolates task-specific information recovery from the subsequent fusion, allowing the same reconciliation mechanism to operate across different deficiencies.

3.3 HIERARCHICAL SCENE–PRIOR FUSION

Given the target-aligned pair (A_q, P_q), our fusion module reconciles their complementary information at two levels: joint interaction in diffusion latent space and prior-guided multi-scale decoding.

Dual-latent interaction. A shared frozen VAE encoder E separately encodes the two predictions:

$$z_q^A = E_z(A_q), \qquad \left(z_q^P, \mathcal{H}_q^P\right) = E(P_q), \qquad \mathcal{H}_q^P = \{h_{q,l}^P\}_{l=1}^L, \tag{3}$$

where z_q^A and z_q^P are the corresponding latent representations and $\mathcal{H}_q^P$ contains multi-scale encoder features from the Prior Prediction. The two latent streams remain separate but interact within the diffusion U-Net. At an attention layer l, let $X_l^A, X_l^P \in \mathbb{R}^{N_l \times d_l}$ denote the spatial tokens of the two streams. We concatenate them along the token dimension and perform self-attention over the joint sequence:

$$X_l = [X_l^A; X_l^P], \qquad \widetilde{X}_l = X_l + \mathrm{softmax}\left(\frac{(X_l W_l^Q)(X_l W_l^K)^\top}{\sqrt{d_l}}\right) X_l W_l^V W_l^O, \tag{4}$$

after which $\widetilde{X}_l$ is split back into the two streams. This joint attention allows each stream to aggregate information from both predictions in a content-dependent manner, rather than performing fixed image blending or one-way feature transfer.

Single-step latent restoration. Conditioned on a fixed diffusion timestep t and restoration instruction embedding e_p, the U-Net U_ϕ predicts noise for both streams ϵ_q^A and ϵ_q^P. We retain the anchor stream and perform one scheduler update:

$$\left(\epsilon_q^A, \epsilon_q^P\right) = U_\phi\left(z_q^A, z_q^P, t, e_p\right), \qquad \widehat{z}_q^A = S_t\left(z_q^A, \epsilon_q^A\right). \tag{5}$$

Although $\widehat{z}_q^A$ corresponds to the anchor stream, ϵ_q^A depends on both inputs through their joint interaction. The restoration therefore remains anchored to the scene-grounded representation while incorporating complementary prior information. The process requires a single U-Net forward pass and scheduler update, without iterative diffusion sampling.

Multi-scale prior-guided decoding. The refined latent is decoded together with the spatial features retained from P_q. At decoder stage l, a matched prior feature is projected by a 1×1 convolution C_l and injected into the decoder:

$$d_{l+1} = \mathrm{UpBlock}_l\left(d_l + C_l\left(h_{q,m(l)}^P\right)\right), \tag{6}$$

where $m(l)$ maps encoder and decoder stages of corresponding spatial resolution. The restored target view is therefore

$$\widehat{I}_q = D_\psi\left(\widehat{z}_q^A; \mathcal{H}_q^P\right). \tag{7}$$

The latent interaction provides a learned reconciliation of the two predictions, while the multi-scale pathway provides an additional spatially aligned route for prior information during decoding. Together, these two levels form the Hierarchical Scene–Prior Fusion module.

For all tasks, the fusion module is supervised by the ground-truth target image $I_q^\star$ using the same restoration objective:

$$\mathcal{L}_{\text{fusion}} = \mathcal{L}_2(\widehat{I}_q, I_q^\star) + \mathcal{L}_{\text{LPIPS-VGG}}(\widehat{I}_q, I_q^\star) + 0.1\,\mathcal{L}_{\text{Gram}}(\widehat{I}_q, I_q^\star). \tag{8}$$

We optimize the full diffusion U-Net, the feature-projection layers used by the decoder skip pathways, and the LoRA parameters of the VAE decoder. The VAE encoder, text encoder, and base VAE decoder parameters remain frozen. A separate fusion checkpoint is trained for each task, while the architecture and objective remain unchanged.

3.4 Restoration-Guided 3D Refinement

The restored views are used to improve the underlying Gaussian representation. Because each $\widehat{I}_q$ retains its target camera q, it can directly provide image supervision for subsequent 3DGS optimization. Let $\widehat{\mathcal{O}} = \{(\widehat{I}_q, q) \mid q \in \mathcal{Q}\}$ denote the restored supervision and let $\widetilde{\mathcal{O}}$ denote the task-dependent supervision set constructed from the available and restored observations. Starting from the scene representation produced by Scene-Grounded Reconstruction, we optimize

$$\mathcal{G}^\star = \arg\min_{\mathcal{G}} \mathcal{L}_{3D}\left(\mathcal{G}; \widetilde{\mathcal{O}}\right), \tag{9}$$

where $\mathcal{L}_{3D}$ is the native reconstruction objective of the underlying Gaussian model. The fusion module remains fixed during this stage. UniFill3D therefore does not introduce a task-specific 3D loss; instead, external prior information is first reconciled with scene-grounded evidence in image space and then consolidated through multi-view 3D optimization.

3.5 Task Instantiations

UniFill3D retains the same target-aligned scene–prior interface, fusion architecture, and 3D refinement principle across all tasks, while the two prediction routes are instantiated with task-appropriate models. Table 1 summarizes the four settings.

Table 1: Task instantiations of UniFill3D.

Task	Anchor	Prior
Sparse view	3DGS (Kerbl et al., 2023)	SEVA (Zhou et al., 2025b)
Super-resolution	Mip-Splatting (Yu et al., 2023)	StableSR (Wang et al., 2024)
Deblurring	Deblurring-3DGS Lee et al. (2024a)	DeblurDiff (Kong et al., 2026)
Low-light	Luminance-GS (Cui et al., 2025)	LightenDiffusion Jiang et al. (2024)

For sparse-view reconstruction, restored views are progressively introduced at additional cameras as pseudo-observations, and a final fusion is applied at query views after 3D refinement. For the three image-degradation tasks, restored input views are generated once at the observed cameras and kept fixed during refinement; novel views are then rendered directly from the refined Gaussian representation. Detailed condition construction and refinement procedures are provided in the Appendix.

4 Experiments

4.1 Experimental Setup

Tasks and datasets. We evaluate UniFill3D on four forms of information deficiency: sparse-view reconstruction, $4\times$ 3D super-resolution, 3D deblurring, and 3D low-light enhancement. For sparse-view reconstruction, we use three input views and evaluate on 12 DL3DV scenes with 526 held-out target views and all nine Mip-NeRF 360 Barron et al. (2022) scenes with 246 target views. For 3D super-resolution, we evaluate on DL3DV (Ling et al., 2024) Benchmark20 (20 scenes, 883 held-out views) and 19 Tanks and Temples scenes (757 held-out views) (Knapitsch et al., 2017). Deblurring is evaluated on ten BlurRF-Synth scenes, each with 29 blurry input views and five held-out sharp views. Low-light enhancement uses four real LOM scenes (Cui et al., 2024) (bike, buu, chair, and sofa) with 16 held-out views. The sparse-view and super-resolution fusion models are trained separately on 970 DL3DV training scenes. The deblurring model is trained on 65 BlurRF-Synth scenes, while the low-light model uses 100 DL3DV scenes with synthetically generated low-light observations. Training and evaluation scenes are disjoint for all tasks.

Evaluation and baselines. All quantitative results are evaluated on held-out target views using PSNR, SSIM, and LPIPS, with FID additionally reported for sparse-view reconstruction. PSNR, SSIM, and LPIPS are first averaged over views within each scene and then averaged equally across

Table 2: **Three-view sparse-view reconstruction.** Comparison on in-domain DL3DV and out-of-domain Mip-NeRF 360. **best** second third

Method	DL3DV (in-domain)				Mip-NeRF 360 (out-of-domain)			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
3DGS (Kerbl et al., 2023)	13.15	0.467	0.613	226.42	12.89	0.242	0.622	244.01
FSGS (Zhu et al., 2024)	14.49	0.558	0.613	180.13	14.26	0.316	0.624	229.29
DepthSplat (Xu et al., 2025)	14.38	0.550	0.544	121.45	15.46	**0.340**	0.617	183.44
SEVA (Zhou et al., 2025b)	15.12	0.529	0.462	46.96	15.08	0.321	0.538	77.90
DiFiX3D+ (Wu et al., 2025a)	14.63	0.541	0.491	54.02	14.01	0.288	0.540	90.52
UniFill3D (Ours)	**16.02**	**0.562**	**0.450**	**46.34**	**15.99**	0.323	**0.524**	**77.47**

Table 3: $4\times$ **3D super-resolution.** Held-out-view evaluation on in-domain DL3DV and out-of-domain Tanks and Temples. **best** second third

Method	DL3DV (in-domain)			Tanks&Temples (out-of-domain)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS (Kerbl et al., 2023)	21.67	0.725	0.349	19.41	0.669	0.350
Mip-Splatting (Yu et al., 2023)	27.24	0.833	0.284	23.10	0.767	0.303
StableSR+3DGS (Wang et al., 2024)	24.31	0.781	0.338	22.47	0.751	0.300
SRGS (Feng et al., 2026)	27.20	0.825	0.283	23.32	0.771	0.286
UniFill3D (Ours)	**27.79**	**0.847**	**0.261**	**23.36**	**0.778**	**0.285**

scenes; FID is computed over all test images pooled within each sparse-view benchmark. We compare against reconstruction-only baselines, task-specific 3D methods, and pipelines that apply external generation or restoration priors before 3D reconstruction. All competing methods are rerun from their public implementations under the same input and evaluation protocol. We denote evaluation on held-out scenes from the fusion-training distribution as *in-domain*, and evaluation on a separate dataset as *out-of-domain*. For sparse-view reconstruction, the reported UniFill3D output includes the final query-view fusion described in Sec. 3.5; the other three tasks are evaluated directly from novel views rendered by the refined Gaussian representation. Held-out target images are used only for evaluation and never as supervision during 3D refinement.

Implementation details. A separate fusion checkpoint is trained for each task using the shared architecture and objective in Sec. 3.3. We use AdamW with a learning rate of 5×10^{-6}, a global batch size of 4, 500 warmup steps, and 10K training iterations. The diffusion timestep is fixed to $t = 199$, with the instruction `remove degradation`. Each fusion model is initialized from the publicly released DiFiX3D+ checkpoint Wu et al. (2025a) and remains frozen during subsequent 3D refinement. Detailed fusion, external-model, and task-specific optimization settings are provided in the supplementary material.

4.2 Evaluation Across Diverse Information Deficiencies in 3D Reconstruction

Incomplete scene coverage: sparse-view reconstruction. Table 2 shows that UniFill3D consistently improves over reconstruction-only, view-generation, and prior-assisted baselines on both DL3DV and Mip-NeRF 360. The gains are especially clear in reconstruction fidelity and perceptual quality, while remaining competitive on structural similarity. Notably, the improvement transfers from DL3DV to the out-of-domain Mip-NeRF 360 benchmark, indicating that the benefit is not restricted to the training distribution. These results support the use of external priors to complement scene-grounded reconstruction when information is missing due to insufficient viewpoint coverage.

Resolution deficiency: 3D super-resolution. As shown in Table 3, UniFill3D achieves the strongest overall performance on both DL3DV and Tanks and Temples, outperforming dedicated 3D super-resolution methods and pipelines that first enhance the input views with an external restoration prior. The consistent improvement across both benchmarks suggests that simply restoring degraded observations before reconstruction is insufficient; jointly reconciling scene-grounded evidence with external prior information provides more effective supervision for the final 3D representation.

Blur degradation: 3D deblurring. Table 4 compares UniFill3D with reconstruction baselines, dedicated 3D deblurring methods, and external deblurring followed by 3D reconstruction. UniFill3D

Table 4: **3D deblurring.** Comparison on the BlurRF-Synth benchmark. **best** second third

Method	PSNR↑	SSIM↑	LPIPS↓
3DGS (Kerbl et al., 2023)	21.50	0.6180	0.4494
DeblurDiff+3DGS (Kong et al., 2026)	24.00	0.6871	0.3627
Deblurring-3DGS (Lee et al., 2024a)	26.93	0.8046	0.2624
BAGS (Peng et al., 2024)	27.49	0.8318	0.2384
PRISM3D (Matta et al., 2026)	24.02	0.7589	0.2609
UniFill3D (Ours)	**28.58**	**0.8634**	**0.1960**

Table 5: **3D low-light enhancement.** Comparison on the real LOM benchmark. **best** second third

Method	PSNR↑	SSIM↑	LPIPS↓
3DGS (Kerbl et al., 2023)	6.50	0.166	0.625
LightenDiffusion+3DGS (Jiang et al., 2024)	16.35	0.702	0.501
LITA-GS (Zhou et al., 2025a)	16.40	0.643	0.544
LL-Gaussian (Sun et al., 2025)	16.26	0.699	0.521
Luminance-GS (Cui et al., 2025)	17.34	0.699	0.515
UniFill3D (Ours)	**18.68**	**0.734**	**0.495**

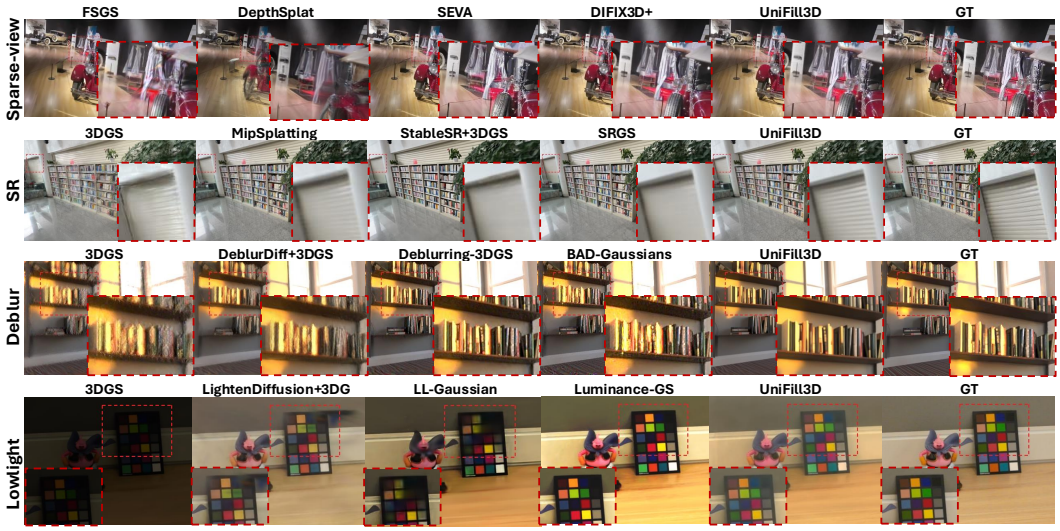


Figure 3: **Qualitative comparison across diverse information deficiencies.** From top to bottom: sparse-view reconstruction, 3D super-resolution, 3D deblurring, and 3D low-light enhancement. Each row compares representative baselines, UniFill3D, and ground truth for the same target view; red dashed insets highlight local details.

performs best across the evaluated fidelity, structural, and perceptual metrics, showing that the same scene–prior formulation extends beyond missing observations to severe image degradation. In particular, the result indicates that external restoration becomes more effective when constrained by the scene-specific information preserved by the 3D reconstruction branch.

Illumination deficiency: 3D low-light enhancement. Table 5 further evaluates UniFill3D under real low-light observations. Despite the substantial domain and appearance shift, UniFill3D achieves the strongest reconstruction fidelity and structural similarity, while remaining competitive in perceptual quality against dedicated low-light methods. This result further supports the generality of the framework: the same mechanism can exploit complementary scene and prior information even when the deficiency arises from illumination rather than missing views, limited resolution, or motion blur.

Qualitative comparison. Figure 3 provides representative visual comparisons across the four information-deficiency settings. UniFill3D consistently yields cleaner and more faithful novel-view renderings than reconstruction-only, task-specific, and prior-assisted alternatives. The improvements are visible in fine-detail recovery for super-resolution, structural sharpness for deblurring, and appearance and visibility under low-light conditions, with similar benefits observed for sparse-view reconstruction. The enlarged regions highlight that UniFill3D better preserves local scene structures rather than merely producing visually enhanced observations, demonstrating that the proposed scene–prior formulation generalizes across markedly different forms of information deficiency.

4.3 Ablations of Scene Grounding and External Priors

We test whether Scene-Grounded Reconstruction and External Prior Recovery genuinely provide complementary information. Table 6 compares independently trained single-branch variants with

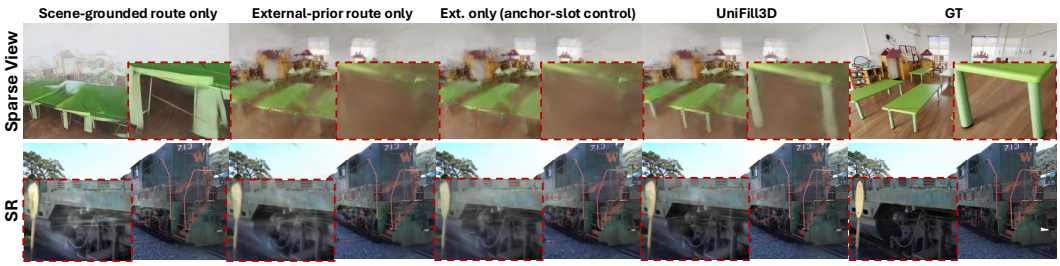


Figure 4: **Qualitative ablations of scene grounding and external priors.** Sparse-view reconstruction (top) and 3D super-resolution (bottom). Columns show the scene-grounded-only model, external-prior-only model, external-prior-only anchor-slot control, UniFill3D, and ground truth. Red dashed insets highlight local details.

the scene–prior model on sparse-view reconstruction and 3D super-resolution. We additionally include an external-prior-only anchor-slot control to account for the asymmetric input roles of the fusion module. The full model performs best across all metrics in both tasks. The relative strength of the two branches varies with the information deficiency: External Prior Recovery is more effective for sparse-view reconstruction, where novel-view information is missing, whereas Scene-Grounded Reconstruction is stronger for super-resolution, where substantial scene evidence remains in the degraded observations. Yet neither branch alone matches their combination. The two external-only slot assignments perform similarly and both remain below the full model, ruling out input-slot placement as the main source of the gain. These results confirm that scene grounding and external priors provide complementary, task-dependent information, and that their joint use is central to UniFill3D.

Figure 4 mirrors the quantitative results. Scene grounding alone preserves scene structure but misses recoverable detail, while external-prior-only variants provide richer appearance with weaker scene fidelity. UniFill3D combines these strengths, producing more faithful local structure and detail in both tasks. The two external-only configurations remain visually similar, further indicating that the improvement comes from scene–prior complementarity rather than input-slot assignment.

Table 6: **Ablations of scene grounding and external priors.** Results on sparse-view reconstruction (DL3DV) and $4\times$ 3D super-resolution (Tanks and Temples). **best** second third

Training variant	DL3DV (3 views)			Tanks and Temples ($4\times$)		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Scene-grounded route only	13.16	0.493	0.563	23.11	0.774	0.289
External-prior route only	15.21	0.526	0.455	22.98	0.747	0.324
External-prior route only (anchor-slot control)	15.34	0.533	0.453	23.02	0.747	0.320
Full scene–prior fusion	**16.02**	**0.562**	**0.450**	**23.36**	**0.778**	**0.285**

5 Conclusion and Limitations

We presented UniFill3D, a unified framework for 3D Gaussian reconstruction under imperfect observations. Rather than treating diverse imperfect-observation settings in isolation, we recast them as manifestations of a common *information-deficiency* problem. UniFill3D addresses this shared challenge by combining scene-grounded reconstruction with external prior recovery through hierarchical scene–prior fusion, and consolidating the restored information through subsequent 3D refinement. Experiments across sparse-view reconstruction, 3D super-resolution, 3D deblurring, and 3D low-light enhancement demonstrate that the same framework achieves strong performance across substantially different observation conditions, while ablations confirm the complementary roles of scene grounding and external priors. These results support information deficiency as a useful common perspective for developing more general 3D reconstruction systems.

The present study covers four representative information-deficiency settings; broader observation conditions and cross-task models would further broaden the empirical scope of UniFill3D.

THE USE OF LARGE LANGUAGE MODELS (LLMS)

Large Language Models were used as a writing assistance tool in this work. Their role was strictly limited to grammar checking, sentence polishing, and improving the fluency and clarity of the manuscript's language. The models did not contribute to research ideas, experimental design, and data analysis. All technical results, claims, and conclusions remain the sole responsibility of the authors.

ETHICS STATEMENT

This work complies with the ICLR Code of Ethics. No ethical concerns requiring specific disclosure are present.

REPRODUCIBILITY STATEMENT

To ensure reproducibility, we provide: (1) implementation details; (2) the training protocols follow the works (Wu et al., 2025a; Zhou et al., 2025b; Feng et al., 2026; Lee et al., 2024a), of which the codes are all released; (3) the training and inference code upon paper acceptance.

REFERENCES

Pranav Asthana, Alex Hanson, Allen Tu, Tom Goldstein, Matthias Zwicker, and Amitabh Varshney. SplatSuRe: Selective Super-Resolution for Multi-view Consistent 3D Gaussian Splatting, 2026. URL https://arxiv.org/abs/2512.02172.

Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5460–5469. IEEE, 2022.

Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *2023 IEEE/CVF international conference on computer vision (ICCV)*, pp. 12470–12479. IEEE, 2023.

Xiao Cao, Yuze Li, Youmin Zhang, Jiayu Song, Cheng Yan, Wen Li, and Lixin Duan. GeoQuery: Geometry-Query Diffusion for Sparse-View Reconstruction, 2026. URL https://arxiv.org/abs/2605.12399.

Eric R. Chan, Koki Nagano, Matthew A. Chan, Alexander W. Bergman, Jeong Joon Park, Axel Levy, Miika Aittala, Shalini De Mello, Tero Karras, and Gordon Wetzstein. Generative Novel View Synthesis with 3D-Aware Diffusion Models, 2023. URL https://arxiv.org/abs/2304.02602.

David Charatan, Sizhe Li, Andrea Tagliasacchi, and Vincent Sitzmann. pixelSplat: 3D Gaussian Splats from Image Pairs for Scalable Generalizable 3D Reconstruction, 2024. URL https://arxiv.org/abs/2312.12337.

Yi-Ting Chen, Ting-Hsuan Liao, Pengsheng Guo, Alexander Schwing, and Jia-Bin Huang. Bridging Diffusion Models and 3D Representations: A 3D Consistent Super-Resolution Framework, 2025. URL https://arxiv.org/abs/2508.04090.

Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. MVSplat: Efficient 3D Gaussian Splatting from Sparse Multi-View Images, 2024. URL https://arxiv.org/abs/2403.14627.

Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-Regularized Optimization for 3D Gaussian Splatting in Few-Shot Images, 2024. URL https://arxiv.org/abs/2311.13398.

Ziteng Cui, Lin Gu, Xiao Sun, Xianzheng Ma, Yu Qiao, and Tatsuya Harada. Aleth-nerf: Illumination adaptive nerf with concealing field assumption. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 1435–1444, 2024.

Ziteng Cui, Xuangeng Chu, and Tatsuya Harada. Luminance-gs: Adapting 3d gaussian splatting to challenging lighting conditions with view-adaptive curve adjustment. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26472–26482. IEEE, 2025.

Xiang Feng, Tieshi Zhong, Shuo Chang, Weiliu Wang, Chengkai Wang, Yifei Chen, Yuhe Wang, Zhenzhong Kuang, Xuefei Yin, and Yanming Zhu. IE-SRGS: An Internal-External Knowledge Fusion Framework for High-Fidelity 3D Gaussian Splatting Super-Resolution, 2025. URL https://arxiv.org/abs/2511.22233.

Xiang Feng, Yongbo He, Linxi Chen, Yan Yang, Chengkai Wang, Yifei Chen, Yixuan Zhong, Zhenzhong Kuang, Jiajun Ding, Xufei Yin, and Yanming Zhu. SRGS: Super-Resolution 3D Gaussian Splatting, 2026. URL https://arxiv.org/abs/2404.10318.

Ruiqi Gao, Aleksander Holynski, Philipp Henzler, Arthur Brussee, Ricardo Martin-Brualla, Pratul Srinivasan, Jonathan T. Barron, and Ben Poole. CAT3D: Create Anything in 3D with Multi-View Diffusion Models, 2024. URL https://arxiv.org/abs/2405.10314.

Yuqi Han, Tao Yu, Xiaohang Yu, Yuwang Wang, and Qionghai Dai. Super-NeRF: View-consistent Detail Generation for NeRF Super-resolution, 2023. URL https://arxiv.org/abs/2304.13518.

Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. Lightendiffusion: Unsupervised low-light image enhancement with latent-retinex diffusion models. In *European conference on computer vision*, pp. 161–179. Springer, 2024.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017.

Lingshun Kong, Jiawei Zhang, Dongqing Zou, Fu Lee Wang, Jimmy S Ren, Xiaohe Wu, Jiangxin Dong, and Jinshan Pan. Deblurdiff: Real-word image deblurring with generative diffusion models. *Advances in Neural Information Processing Systems*, 38:154855–154873, 2026.

Byeonghyeon Lee, Howoong Lee, Xiangyu Sun, Usman Ali, and Eunbyung Park. Deblurring 3d gaussian splatting. In *European Conference on Computer Vision*, pp. 127–143. Springer, 2024a.

Jie Long Lee, Chen Li, and Gim Hee Lee. DiSR-NeRF: Diffusion-Guided View-Consistent Super-Resolution NeRF, 2024b. URL https://arxiv.org/abs/2404.00874.

Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. DNGaussian: Optimizing Sparse-View 3D Gaussian Radiance Fields with Global-Local Depth Normalization, 2024. URL https://arxiv.org/abs/2403.06912.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior, 2024. URL https://arxiv.org/abs/2308.15070.

Lu Ling, Yichen Sheng, Zhi Tu, Wentian Zhao, Cheng Xin, Kun Wan, Lantao Yu, Qianyu Guo, Zixun Yu, Yawen Lu, et al. Dl3dv-10k: A large-scale scene dataset for deep learning-based 3d vision. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22160–22169. IEEE, 2024.

Fangfu Liu, Wenqiang Sun, Hanyang Wang, Yikai Wang, Haowen Sun, Junliang Ye, Jun Zhang, and Yueqi Duan. ReconX: Reconstruct Any Scene from Sparse Views with Video Diffusion Model, 2025. URL https://arxiv.org/abs/2408.16767.

Xi Liu, Chaoyi Zhou, and Siyu Huang. 3DGS-Enhancer: Enhancing Unbounded 3D Gaussian Splatting with View-consistent 2D Diffusion Priors, 2024a. URL https://arxiv.org/abs/2410.16266.

Xinhang Liu, Jiaben Chen, Shiu-Hong Kao, Yu-Wing Tai, and Chi-Keung Tang. Deceptive-NeRF/3DGS: Diffusion-Generated Pseudo-Observations for High-Quality Sparse-View Reconstruction, 2024b. URL https://arxiv.org/abs/2305.15171.

Gopi Raju Matta, Trisha Reddypalli, Divya Madhuri Vemunuri, and Kaushik Mitra. Prism3d: Probabilistic refinement and robust initialization for physically consistent scene modeling under extreme motion blur. In *European Conference on Computer Vision*, pp. 167–186. Springer, 2026.

Michael Niemeyer, Jonathan T. Barron, Ben Mildenhall, Mehdi S. M. Sajjadi, Andreas Geiger, and Noha Radwan. RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs, 2021. URL https://arxiv.org/abs/2112.00724.

Cheng Peng, Yutao Tang, Yifan Zhou, Nengyu Wang, Xijun Liu, Deming Li, and Rama Chellappa. Bags: Blur agnostic gaussian splatting through multi-scale kernel modeling. In *European Conference on Computer Vision*, pp. 293–310. Springer, 2024.

Kyle Sargent, Zizhang Li, Tanmay Shah, Charles Herrmann, Hong-Xing Yu, Yunzhi Zhang, Eric Ryan Chan, Dmitry Lagun, Li Fei-Fei, Deqing Sun, and Jiajun Wu. ZeroNVS: Zero-Shot 360-Degree View Synthesis from a Single Image, 2024. URL https://arxiv.org/abs/2310.17994.

Yuan Shen, Duygu Ceylan, Paul Guerrero, Zexiang Xu, Niloy J. Mitra, Shenlong Wang, and Anna Frühstück. SuperGaussian: Repurposing Video Models for 3D Super Resolution, 2024. URL https://arxiv.org/abs/2406.00609.

Hao Sun, Fenggen Yu, Huiyao Xu, Tao Zhang, and Changqing Zou. Ll-gaussian: Low-light scene reconstruction and enhancement via gaussian splatting for novel view synthesis. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 4261–4270, 2025.

Yecong Wan, Mingwen Shao, Yuanshuo Cheng, and Wangmeng Zuo. S2Gaussian: Sparse-View Super-Resolution 3D Gaussian Splatting, 2025. URL https://arxiv.org/abs/2503.04314.

Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. NeRF-SR: High-Quality Neural Radiance Fields using Supersampling, 2022. URL https://arxiv.org/abs/2112.01759.

Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting Diffusion Prior for Real-World Image Super-Resolution, 2024. URL https://arxiv.org/abs/2305.07015.

Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, and Bihan Wen. SinSR: Diffusion-Based Image Super-Resolution in a Single Step, 2023. URL https://arxiv.org/abs/2311.14760.

Frederik Warburg, Ethan Weber, Matthew Tancik, Aleksander Holynski, and Angjoo Kanazawa. Nerfbusters: Removing Ghostly Artifacts from Casually Captured NeRFs, 2023. URL https://arxiv.org/abs/2304.10532.

Daniel Watson, William Chan, Ricardo Martin-Brualla, Jonathan Ho, Andrea Tagliasacchi, and Mohammad Norouzi. Novel View Synthesis with Diffusion Models, 2022. URL https://arxiv.org/abs/2210.04628.

Jay Zhangjie Wu, Yuxuan Zhang, Haithem Turki, Xuanchi Ren, Jun Gao, Mike Zheng Shou, Sanja Fidler, Zan Gojcic, and Huan Ling. DiFiX3D+: Improving 3D Reconstructions with Single-Step Diffusion Models, 2025a. URL https://arxiv.org/abs/2503.01774.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-Step Effective Diffusion Network for Real-World Image Super-Resolution, 2024a. URL https://arxiv.org/abs/2406.08177.

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution, 2024b. URL https://arxiv.org/abs/2311.16518.

Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P. Srinivasan, Dor Verbin, Jonathan T. Barron, Ben Poole, and Aleksander Holynski. ReconFusion: 3D Reconstruction with Diffusion Priors, 2023. URL https://arxiv.org/abs/2312.02981.

Sibo Wu, Congrong Xu, Binbin Huang, Andreas Geiger, and Anpei Chen. GenFusion: Closing the Loop between Reconstruction and Generation via Videos, 2025b. URL https://arxiv.org/abs/2503.21219.

Jamie Wynn and Daniyar Turmukhambetov. DiffusioNeRF: Regularizing Neural Radiance Fields with Denoising Diffusion Models, 2023. URL https://arxiv.org/abs/2302.12231.

Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13049–13059. IEEE, 2023.

Haofei Xu, Songyou Peng, Fangjinhua Wang, Hermann Blum, Daniel Barath, Andreas Geiger, and Marc Pollefeys. DepthSplat: Connecting Gaussian Splatting and Depth, 2025. URL https://arxiv.org/abs/2410.13862.

Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-Aware Stable Diffusion for Realistic Image Super-resolution and Personalized Stylization, 2024. URL https://arxiv.org/abs/2308.14469.

Xingyilang Yin, Qi Zhang, Jiahao Chang, Ying Feng, Qingnan Fan, Xi Yang, Chi-Man Pun, Huaqi Zhang, and Xiaodong Cun. GSFixer: Improving 3D Gaussian Splatting with Reference-Guided Video Diffusion Priors, 2025. URL https://arxiv.org/abs/2508.09667.

Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild, 2024a. URL https://arxiv.org/abs/2401.13627.

Xiqian Yu, Hanxin Zhu, Tianyu He, and Zhibo Chen. GaussianSR: 3D Gaussian Super-Resolution with 2D Diffusion Priors, 2024b. URL https://arxiv.org/abs/2406.10111.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-Splatting: Alias-free 3D Gaussian Splatting, 2023. URL https://arxiv.org/abs/2311.16493.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739, 2022.

Jiawei Zhang, Jiahe Li, Xiaohan Yu, Lei Huang, Lin Gu, Jin Zheng, and Xiao Bai. CoR-GS: Sparse-View 3D Gaussian Splatting via Co-Regularization, 2024. URL https://arxiv.org/abs/2405.12110.

Lingzhe Zhao, Peng Wang, and Peidong Liu. Bad-gaussians: Bundle adjusted deblur gaussian splatting. In *European Conference on Computer Vision*, pp. 233–250. Springer, 2024.

Han Zhou, Wei Dong, and Jun Chen. Lita-gs: Illumination-agnostic novel view synthesis via reference-free 3d gaussian splatting and physical priors. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21580–21589. IEEE, 2025a.

Jensen Zhou, Hang Gao, Vikram Voleti, Aaryaman Vasishta, Chun-Han Yao, Mark Boss, Philip Torr, Christian Rupprecht, and Varun Jampani. Stable Virtual Camera: Generative View Synthesis with Diffusion Models, 2025b. URL https://arxiv.org/abs/2503.14489.

Zhizhuo Zhou and Shubham Tulsiani. SparseFusion: Distilling View-conditioned Diffusion for 3D Reconstruction, 2023. URL https://arxiv.org/abs/2212.00792.

Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. FSGS: Real-Time Few-shot View Synthesis using Gaussian Splatting, 2024. URL https://arxiv.org/abs/2312.00451.

A Additional Method Details

A.1 Fusion Implementation

This section provides implementation details of the Hierarchical Scene–Prior Fusion module described in Sec. 3.3. For a target-aligned pair (A_q, P_q), both images are encoded independently by the shared VAE encoder. Latents are sampled from the VAE posterior rather than using the posterior mean. Following encoding, the sampled latents are multiplied by the scaling factor 0.18215; the inverse scaling is applied before VAE decoding.

The two latent streams jointly pass through the diffusion U-Net and interact through the joint self-attention described in the main paper. The U-Net predicts

$$(\epsilon_q^A, \epsilon_q^P) = U_\phi(z_q^A, z_q^P, t, e_p).$$

Both branches undergo the scheduler update, while the updated anchor branch is retained for final decoding. We use a `DDPMScheduler` with 1,000 training timesteps, a `scaled_linear` beta schedule from 0.00085 to 0.012, ϵ-prediction, and `clip_sample=false`. Inference explicitly sets

$$\texttt{timesteps} = [199],$$

so the fusion process performs one U-Net evaluation and one scheduler update rather than 199 iterative denoising steps.

The updated anchor latent is decoded together with multi-scale features extracted from the Prior Prediction. Four feature injections are used, each projected by a 1×1 convolution and added before the corresponding decoder upsampling block. The exact feature dimensions are summarized in Table 7.

Table 7: Multi-scale prior-feature injection in the VAE decoder. Spatial sizes are relative to the output image resolution.

Injection location	Spatial resolution	Prior $\rightarrow$ decoder channels
Before `up_blocks[0]`	$H/8 \times W/8$	$512 \rightarrow 512$
Before `up_blocks[1]`	$H/4 \times W/4$	$256 \rightarrow 512$
Before `up_blocks[2]`	$H/2 \times W/2$	$128 \rightarrow 512$
Before `up_blocks[3]`	$H \times W$	$128 \rightarrow 256$

The full diffusion U-Net, the four feature-projection convolutions, and the LoRA parameters of the VAE decoder are optimized during fusion training. The VAE encoder, text encoder, and base parameters of the VAE decoder remain frozen. The same architecture is used for all four tasks; separate task-specific checkpoints are trained with the common objective defined in Sec. 3.3.

A.2 Construction of Target-Aligned Training Pairs

Fusion training requires a Reconstruction Anchor A_q, a Prior Prediction P_q, and the ground-truth target image $I_q^\star$ at the same camera q. Across all tasks, the two predictions are generated independently from the deficient observations; in particular, the external model does not take the internal reconstruction rendering as its input.

For sparse-view reconstruction, a 3DGS representation is optimized from three input views, with SfM initialization restricted to points visible from those views. At a held-out target camera q, the 30K-step 3DGS renders A_q, while SEVA predicts P_q from the same three input views at camera q. The corresponding held-out image provides $I_q^\star$ during fusion training.

For $4\times$ 3D super-resolution, Mip-Splatting reconstructs the scene from low-resolution observations and renders A_q on the high-resolution image plane of an observed camera. StableSR-Turbo independently super-resolves the corresponding low-resolution observation to produce P_q. The original high-resolution image provides the training target.

For 3D deblurring, Deblurring-3DGS is optimized from the blurry observations and its 20K-step sharp rendering at an observed camera forms A_q. DeblurDiff independently processes the corre-

sponding blurry observation to produce P_q, while the sharp image at that camera provides supervision.

For 3D low-light enhancement, Luminance-GS is reconstructed from low-light observations and its 10K-step output at an observed camera forms A_q. LightenDiffusion independently enhances the corresponding low-light observation to produce P_q, with the clean image providing the target.

Thus, the four tasks differ in how their two predictions are generated, but every fusion sample has the same form,

$$(A_q, P_q, I_q^\star),$$

with all three images defined at the same target camera and image plane.

B Additional Experimental Details

B.1 Evaluation Protocols

Unless otherwise specified, PSNR, SSIM, and LPIPS are computed for each held-out target view, averaged within each scene, and then averaged equally across scenes. This scene-balanced aggregation prevents scenes containing more target views from dominating the reported results. LPIPS uses the VGG backbone throughout. For sparse-view reconstruction, FID is additionally computed over all predicted and ground-truth test images pooled within each evaluation dataset.

For sparse-view reconstruction, DL3DV is evaluated at 1024×576, while Mip-NeRF 360 uses the native resolution of each scene. Super-resolution is evaluated at the target high resolution. Deblurring is evaluated at 600×400, and low-light enhancement is evaluated on 448×448 center crops. All competing methods within each benchmark are evaluated using the same target views, image domain, and metric implementation.

Held-out target images are used exclusively for evaluation and are never used as supervision during restoration-guided 3D refinement.

B.2 Fusion Training

A separate fusion model is trained for each task using the same architecture and restoration objective in Sec. 3.3. Each model is initialized from the publicly released DiFiX3D+ checkpoint Wu et al. (2025a). We optimize using AdamW with a learning rate of 5×10^{-6}, a global batch size of 4, and 500 warmup steps. Training runs for 10K optimization steps.

The diffusion timestep is fixed to $t = 199$, and the text instruction is `remove degradation`. All four tasks use

$$\mathcal{L}_{\text{fusion}} = \mathcal{L}_2 + \mathcal{L}_{\text{LPIPS-VGG}} + 0.1\,\mathcal{L}_{\text{Gram}}.$$

The trainable and frozen parameter groups are detailed in Sec. A.1. The fusion model is frozen during all subsequent 3D refinement.

B.3 External Prior Model Configurations

We use the publicly released checkpoints of all external prior models. Their inference configurations are summarized in Table 8.

Table 8: Inference configurations of the external prior models.

Model	Configuration
SEVA	v1.1; 50 inference steps; CFG $= 2.0$; $\text{CFG}_{\min} = 1.2$; seed 23.
StableSR-Turbo	`stablesr_turbo.ckpt` with `vqgan_cfw_00011.ckpt`; 4 inference steps; dec_w $= 0.5$; seed 42; wavelet color correction; $4\times$ upsampling.
DeblurDiff	Released `model.pth`; 50 inference steps; seed 231; `better_start=false`.
LightenDiffusion	Released stage-1 and stage-2 weights; $T = 1000$; $S = 20$; seed 0; `ema=false`.

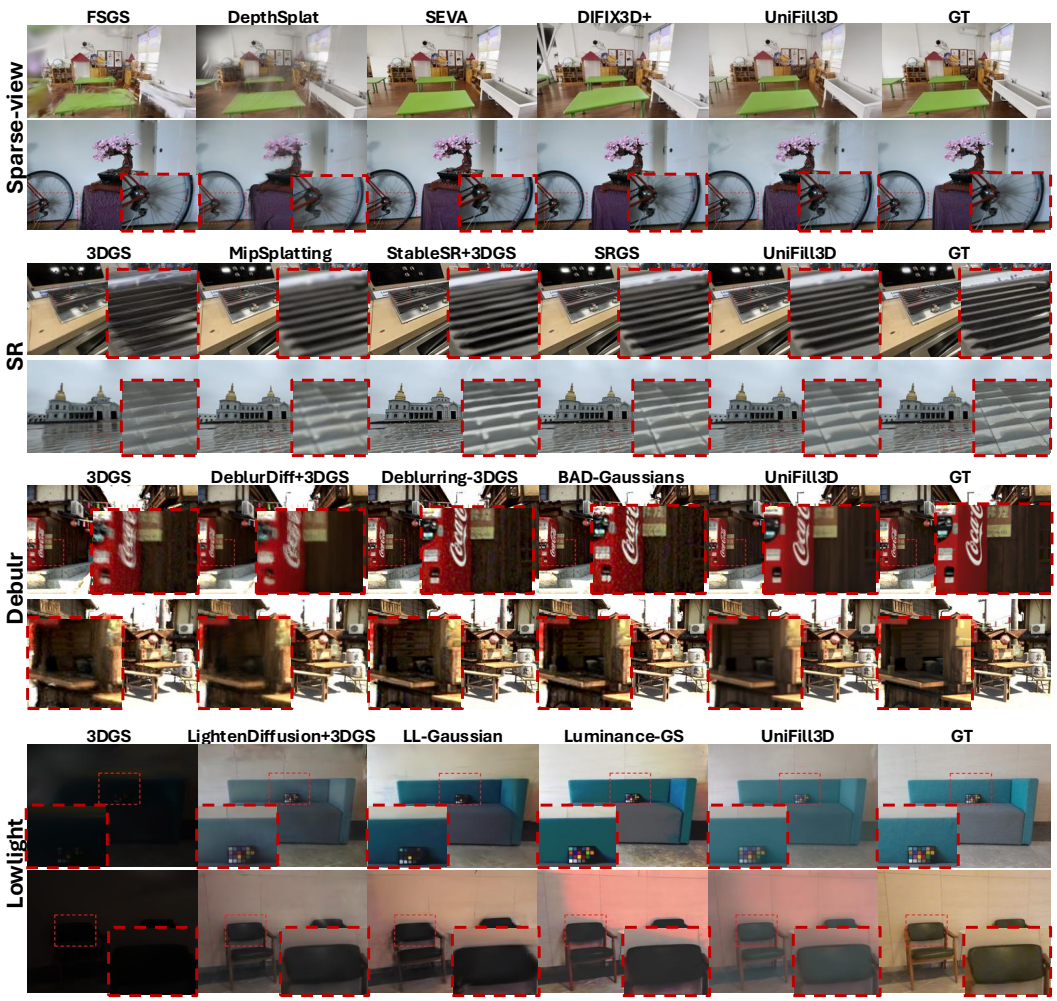


Figure 5: **Additional qualitative comparison across diverse information deficiencies.** From top to bottom: sparse-view reconstruction, 3D super-resolution, 3D deblurring, and 3D low-light enhancement. Each row compares representative baselines, UniFill3D, and ground truth for the same target view; red dashed insets highlight local details.

B.4 Synthetic Low-Light Training Data

The low-light fusion model is trained on 100 DL3DV scenes with synthetic low-light observations generated following the degradation process used in Lumos3D. Given a clean sRGB image I_{clean}, we construct

$$I_{\text{low}} = \text{sRGB}\left[(e \cdot \text{Linear}(I_{\text{clean}}))^{\gamma}\right], \tag{10}$$

where

$$e \sim \mathcal{U}(0.05, 0.10), \qquad \gamma \sim \mathcal{U}(1.30, 1.40).$$

A single (e, γ) pair is sampled for each scene and shared across all views of that scene, preserving cross-view consistency in the synthetic degradation. Parameters are generated with seed 20260831. We do not introduce additional white-balance perturbation or sensor noise. The degradation is applied to the original-resolution images before resizing them to 1024×576 for training.

B.5 Additional Qualitative Results

Figure 5 provides additional qualitative comparisons across all four information-deficiency settings. The results are consistent with the observations in the main paper. For sparse-view reconstruction,

UniFill3D better preserves scene structure and appearance at novel viewpoints, with fewer distortions in locally challenging regions. For 3D super-resolution, it recovers sharper edges and finer repetitive structures while remaining faithful to the underlying scene. In deblurring, UniFill3D produces clearer local structures and text-like details with reduced residual blur. Under low-light observations, it restores visibility and appearance while better preserving scene content than the competing approaches. Across these examples, the improvements are not limited to a particular degradation type or scene. The enlarged regions further illustrate that UniFill3D consistently balances scene fidelity with detail recovery, supporting the complementary use of scene-grounded reconstruction and external priors.

B.6 Per-Scene Quantitative Results

Table 9: **Per-scene $4\times$ 3D super-resolution on DL3DV Benchmark20.** 20 scenes and 883 held-out views. Scene IDs are abbreviated to their first eight characters.

Scene	3DGS			Mip-Splatting			StableSR + 3DGS			SRGS			**UniFill3D (Ours)**		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
032dee9f	20.22	0.670	0.412	28.44	0.834	0.314	25.28	0.785	0.347	28.42	0.824	0.310	**29.51**	**0.860**	**0.266**
08539793	23.61	0.793	0.303	30.48	0.891	0.243	26.93	0.838	0.304	30.39	0.879	0.249	**31.18**	**0.903**	**0.221**
093ef327	21.75	0.777	0.323	27.40	0.873	0.253	24.79	0.826	0.317	27.01	0.861	0.269	**27.90**	**0.887**	**0.233**
1d6a9ed4	17.88	0.534	0.430	**22.44**	**0.688**	**0.347**	20.53	0.622	0.420	22.19	0.671	0.356	22.24	0.678	0.361
26fd2335	20.96	0.621	0.400	26.09	0.736	0.331	23.84	0.671	0.402	25.69	0.711	0.361	**26.49**	**0.752**	**0.319**
280abf7b	23.54	0.798	0.303	27.78	0.873	0.266	24.69	0.812	0.337	27.86	0.866	0.266	**28.14**	**0.881**	**0.245**
2beaca31	21.59	0.746	0.293	**26.93**	**0.850**	**0.229**	23.36	0.763	0.311	25.97	0.812	0.254	26.49	0.838	0.231
341b4ff3	20.99	0.748	0.334	25.13	0.837	0.277	22.46	0.785	0.322	25.25	0.829	0.273	**25.71**	**0.860**	**0.242**
374ffd0c	24.19	0.782	0.317	30.47	0.881	0.251	26.96	0.836	0.290	30.31	0.875	0.243	**31.32**	**0.904**	**0.200**
41036716	24.10	0.820	0.290	29.18	0.896	0.239	26.19	0.859	0.285	29.25	0.891	0.234	**29.92**	**0.910**	**0.213**
484c0aca	21.82	0.795	0.338	26.63	0.877	0.294	22.47	0.833	0.329	27.16	0.893	0.249	**27.47**	**0.904**	**0.238**
4ff8650b	23.64	0.813	0.288	29.79	0.903	0.229	26.93	0.868	0.273	29.95	0.898	0.226	**30.73**	**0.916**	**0.206**
513e4ea2	17.83	0.604	0.459	26.64	0.811	0.319	22.67	0.755	0.377	26.21	0.803	0.313	**27.34**	**0.838**	**0.286**
8b9fb9d9	18.29	0.580	0.414	**23.52**	**0.737**	**0.327**	20.85	0.654	0.409	23.21	0.723	0.334	23.36	0.730	0.337
91afb991	22.97	0.783	0.313	28.50	0.871	0.251	24.88	0.825	0.300	28.34	0.865	0.247	**29.19**	**0.885**	**0.224**
ba55c875	22.78	0.778	0.326	27.96	0.861	0.278	25.08	0.819	0.338	27.64	0.852	0.284	**28.42**	**0.869**	**0.266**
c076929d	22.04	0.759	0.339	26.08	0.825	0.330	25.12	0.809	0.324	**28.57**	**0.855**	**0.263**	26.99	0.850	0.277
ceb252f5	23.18	0.777	0.325	29.10	0.875	0.259	26.46	0.843	0.287	29.10	0.873	0.250	**30.02**	**0.895**	**0.228**
df29c225	20.89	0.649	0.384	25.43	0.755	0.308	22.97	0.668	0.399	24.75	0.717	0.337	**25.92**	**0.764**	**0.305**
ff592398	21.11	0.675	0.397	26.74	0.791	0.337	23.66	0.745	0.381	26.70	0.793	0.333	**27.48**	**0.818**	**0.318**
Mean	21.67	0.725	0.349	27.24	0.833	0.284	24.31	0.781	0.338	27.20	0.825	0.283	**27.79**	**0.847**	**0.261**

Tables 9–11 report per-scene results for 3D super-resolution, 3D deblurring, and 3D low-light enhancement. All results are computed on the same held-out target views used in the corresponding main-paper evaluations. Metrics are averaged over target views within each scene, while the mean row gives the scene-balanced average reported in the main tables. The best result among the compared methods is highlighted in bold.

B.6.1 3D Super-Resolution

Tables 9 and 10 provide per-scene results on DL3DV and Tanks and Temples, respectively. UniFill3D achieves the strongest scene-averaged performance on both datasets across PSNR, SSIM, and LPIPS. The gains are also distributed across a broad range of individual scenes rather than being driven by a small subset, although task-specific baselines remain stronger on some scenes. This trend is consistent on both the in-domain DL3DV benchmark and the out-of-domain Tanks and Temples benchmark.

Table 11 reports results on the four real LOM scenes. UniFill3D achieves the best scene-averaged performance for all three metrics and performs strongly across scenes with substantially different appearance and illumination. The per-scene comparison further shows that no single competing method dominates across all scenes or metrics, whereas UniFill3D provides the strongest overall aggregate performance.

Table 10: **Per-scene** $4\times$ **3D super-resolution on Tanks and Temples.** 19 scenes and 757 held-out views.

Scene	3DGS			Mip-Splatting			StableSR + 3DGS			SRGS			**UniFill3D (Ours)**		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
auditorium	22.14	0.807	0.301	23.46	0.838	0.275	23.62	0.849	0.247	**24.23**	**0.861**	**0.242**	23.78	0.857	0.249
ballroom	18.23	0.578	0.337	23.15	0.729	0.288	22.20	0.702	0.283	23.44	0.734	0.266	**23.51**	**0.745**	**0.261**
barn	22.97	0.729	0.317	26.85	0.790	0.293	25.18	0.771	0.295	26.83	0.794	0.277	**27.29**	**0.810**	**0.262**
caterpillar	19.36	0.602	0.385	22.99	0.704	0.325	22.44	0.686	0.340	22.98	0.703	**0.321**	**23.10**	**0.710**	0.329
church	20.00	0.682	0.343	21.99	0.753	0.313	22.02	0.744	0.306	**22.54**	0.761	**0.294**	22.42	**0.763**	0.297
courtroom	19.53	0.647	0.376	22.26	0.730	0.323	21.75	0.725	0.294	**22.74**	0.746	0.292	22.62	**0.751**	**0.285**
family	15.09	0.566	0.408	24.09	0.788	0.281	22.54	0.753	0.289	23.82	0.780	0.271	**24.24**	**0.799**	**0.252**
francis	22.47	0.779	0.352	26.62	0.855	0.316	25.97	0.845	0.316	26.55	0.857	**0.304**	**27.15**	**0.866**	**0.304**
horse	16.30	0.675	0.322	23.48	0.827	0.237	22.72	0.811	0.234	23.72	0.829	0.221	**23.78**	**0.837**	**0.215**
ignatius	17.13	0.527	0.411	21.22	**0.681**	0.331	20.52	0.646	0.357	21.26	0.671	**0.328**	**21.27**	0.676	0.347
lighthouse	20.47	0.742	0.303	21.29	0.789	0.296	21.65	0.777	0.283	21.74	0.789	0.276	**22.06**	**0.799**	**0.266**
m60	21.72	0.735	0.302	26.16	0.836	0.250	24.72	0.810	0.266	26.17	0.835	0.242	**26.59**	**0.848**	**0.229**
museum	17.44	0.579	0.400	20.28	0.689	0.345	19.76	0.683	0.317	**20.63**	0.704	**0.315**	20.41	**0.709**	0.316
palace	18.35	0.666	0.380	18.24	0.687	0.408	19.68	0.709	**0.364**	**19.69**	**0.713**	0.367	18.36	0.694	0.397
panther	22.49	0.752	0.309	27.11	0.847	0.260	25.23	0.819	0.275	26.98	0.846	0.251	**27.46**	**0.860**	**0.233**
playground	19.85	0.628	0.371	24.13	0.746	0.316	23.24	0.720	0.333	24.12	0.742	0.311	**24.29**	**0.756**	**0.306**
temple	18.99	0.715	0.335	20.31	0.762	0.325	20.24	0.761	0.302	**20.49**	**0.773**	**0.299**	19.92	0.772	0.303
train	18.80	0.648	0.346	21.33	0.733	0.298	20.55	0.705	0.305	21.28	0.729	0.289	**21.49**	**0.741**	**0.288**
truck	17.38	0.649	0.345	23.94	0.786	0.284	22.86	0.759	0.298	23.87	0.784	0.274	**24.18**	**0.796**	**0.269**
Mean	19.41	0.669	0.350	23.10	0.767	0.303	22.47	0.751	0.300	23.32	0.771	0.286	**23.36**	**0.778**	**0.285**

Table 11: **Per-scene 3D low-light enhancement on LOM.** 4 scenes and 16 held-out views. All methods are evaluated on the same four scenes and 16 held-out views.

Scene	3DGS			LightenDiffusion + 3DGS			LITA-GS			Luminance-GS			**UniFill3D (Ours)**		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
bike	6.17	0.059	0.682	14.30	0.597	0.494	**19.51**	**0.673**	**0.375**	16.39	0.627	0.520	16.49	0.628	0.484
buu	7.27	0.267	0.516	**21.61**	0.777	0.416	9.40	0.475	0.828	15.40	0.725	0.436	19.91	**0.778**	**0.402**
chair	6.09	0.140	0.651	14.98	0.693	0.590	18.74	0.694	**0.537**	18.58	0.690	0.634	**18.92**	**0.729**	0.630
sofa	6.46	0.200	0.652	14.53	0.741	0.503	17.97	0.729	**0.437**	18.98	0.756	0.472	**19.40**	**0.801**	0.465
Mean	6.50	0.166	0.625	16.35	0.702	0.501	16.40	0.643	0.544	17.34	0.699	0.515	**18.68**	**0.734**	**0.495**

B.6.2 3D DEBLURRING

Table 12 reports results on the ten BlurRF-Synth test scenes. UniFill3D obtains the best scene-averaged PSNR, SSIM, and LPIPS, with improvements observed across most individual scenes. The per-scene results also show that the advantage is not uniform: dedicated deblurring baselines remain competitive on several scenes, while UniFill3D provides the strongest overall balance across reconstruction fidelity, structural similarity, and perceptual quality.

Table 12: **Per-scene 3D deblurring on BlurRF-Synth Test10.** Ten scenes, each with five held-out sharp target views.

Scene	3DGS	DeblurDiff + 3DGS	Deblurring-3DGS	BAGS	PRISM3D	**UniFill3D (Ours)**
			PSNR↑			
cozyroom	22.65	23.78	28.18	27.87	25.02	**28.54**
entrance	24.59	25.79	29.17	27.51	25.08	**30.58**
factory	18.51	18.60	23.49	23.82	21.27	**24.21**
homelibrary_1	25.72	26.70	28.86	29.44	24.83	**31.28**
homelibrary_2	24.88	24.87	26.93	27.26	24.94	**29.87**
interior	14.03	27.55	31.22	32.20	27.47	**33.56**
lotus	26.54	26.39	27.93	29.50	27.15	**30.49**
pool	19.71	27.67	28.28	28.65	27.06	**29.80**
tanabata	19.19	19.50	22.33	**23.96**	18.31	23.95
wine	19.23	19.11	22.92	**24.67**	19.05	23.56
Mean	21.50	24.00	26.93	27.49	24.02	**28.58**
			SSIM↑			
cozyroom	0.7334	0.7706	0.8724	0.8857	0.8232	**0.9072**
entrance	0.7598	0.8072	0.8603	0.8746	0.8432	**0.9287**
factory	0.5757	0.5362	0.7929	**0.8008**	0.7693	0.7926
homelibrary_1	0.6840	0.6983	0.7944	0.7939	0.7139	**0.8801**
homelibrary_2	0.6511	0.6412	0.7468	0.7777	0.7281	**0.8452**
interior	0.3087	0.8112	0.8858	0.9193	0.8636	**0.9451**
lotus	0.6766	0.6341	0.7472	0.8388	0.7752	**0.8450**
pool	0.4642	0.7066	0.7904	0.7972	0.7518	**0.8392**
tanabata	0.6555	0.6406	0.7511	0.7911	0.6314	**0.8285**
wine	0.6713	0.6245	0.8050	**0.8390**	0.6891	0.8219
Mean	0.6180	0.6871	0.8046	0.8318	0.7589	**0.8634**
			LPIPS-VGG↓			
cozyroom	0.3623	0.2806	0.1838	0.1913	0.1980	**0.1357**
entrance	0.3644	0.2871	0.2321	0.1943	0.1684	**0.1036**
factory	0.4499	0.4508	0.2874	0.3119	0.3070	**0.2801**
homelibrary_1	0.4332	0.3595	0.2910	0.3011	0.3012	**0.1685**
homelibrary_2	0.4000	0.3658	0.2621	0.2325	0.2548	**0.2064**
interior	0.7045	0.3029	0.2233	0.1686	0.1622	**0.1078**
lotus	0.4474	0.4322	0.3525	0.2687	0.2828	**0.2436**
pool	0.5724	0.3835	0.2959	0.2721	0.3006	**0.2468**
tanabata	0.3993	0.3982	0.2784	0.2507	0.3354	**0.2476**
wine	0.3611	0.3665	0.2178	**0.1926**	0.2987	0.2198
Mean	0.4494	0.3627	0.2624	0.2384	0.2609	**0.1960**