Beyond Darkness: Thermal-Supervised 3D Gaussian Splatting for Low-Light Novel View Synthesis

Qingsen Ma*   Chen Zou*   Dianyun Wang*   Jia Wang*   Liuyu Xiang   Zhaofeng He†

Beijing University of Posts and Telecommunications

maqingsen@bupt.edu.cn

Abstract

Under extremely low-light conditions, novel view synthesis (NVS) faces severe degradation in terms of geometry, color consistency, and radiometric stability. Standard 3D Gaussian Splatting (3DGS) pipelines fail when applied directly to underexposed inputs, as independent enhancement across views causes illumination inconsistencies and geometric distortion. To address this, we present DTGS, a unified framework that tightly couples Retinex-inspired illumination decomposition with thermal-guided 3D Gaussian Splatting for illumination-invariant reconstruction. Unlike prior approaches that treat enhancement as a pre-processing step, DTGS performs joint optimization across enhancement, geometry, and thermal supervision through a cyclic enhancement–reconstruction mechanism. A thermal supervisory branch stabilizes both color restoration and geometry learning by dynamically balancing enhancement, structural, and thermal losses. Moreover, a Retinex-based decomposition module embedded within the 3DGS loop provides physically interpretable reflectance–illumination separation, ensuring consistent color and texture across viewpoints. To evaluate our method, we construct RGBT-LOW, a new multi-view low-light thermal dataset capturing severe illumination degradation. Extensive experiments show that DTGS significantly outperforms existing low-light enhancement and 3D reconstruction baselines, achieving superior radiometric consistency, geometric fidelity, and color stability under extreme illumination.

1. Introduction

Novel View Synthesis (NVS) is a fundamental task in computer vision with broad applications in augmented and virtual reality (AR/VR) [3, 8, 10, 15]. Recent advances such as Neural Radiance Fields (NeRF) [17] and 3D Gaussian Splatting (3DGS) [12] have achieved high-quality render-

---

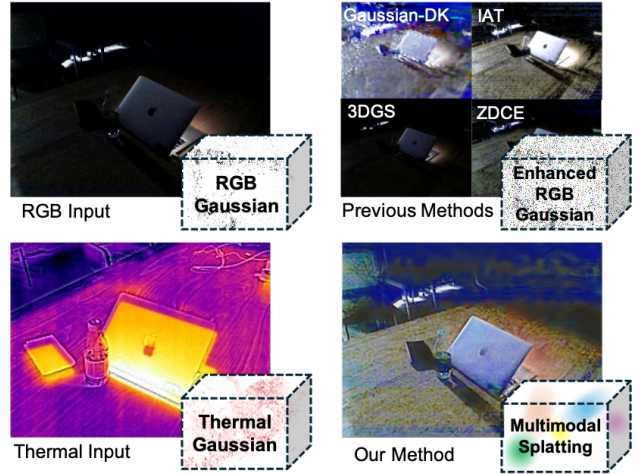*These authors contributed equally.
†Corresponding author.

Figure 1. RGB-only enhancement struggles to maintain structural coherence and color consistency in extremely low-light scenes. Our method delivers more consistent colors and superior contrast.

ing and impressive efficiency, enabling real-time or near real-time performance. However, these successes typically rely on multi-view images that are well-exposed and of high signal-to-noise ratio. In real-world scenarios—e.g., archaeological exploration in dark caves, nighttime military simulation, nocturnal driving, or robotic search-and-rescue in poorly lit environments—illumination is harsh or extremely limited, and existing NVS pipelines [18, 24] struggle to preserve color consistency, structural fidelity, and geometric interpretability.

Directly applying 3DGS to images captured under adverse illumination is challenging for several reasons. First, severe under-exposure, significant noise, and color distortion cause substantial information loss, making it difficult to recover high-fidelity geometry and radiometry in the absence of ground-truth references. Second, naively combining exposure-correction or low-light enhancement [9, 14, 33] with 3DGS often introduces cross-view illumination inconsistencies: each view is enhanced independently,

arXiv:2511.13011v1 [cs.CV] 17 Nov 2025

so inter-view lighting statistics diverge, degrading multi-view consistency and reconstruction quality. Third, extreme darkness exacerbates color shifts and structural blurring; even state-of-the-art enhancement models [1, 21, 30]can produce color artifacts, hue drift, or over-smoothing that harm downstream 3D reconstruction.

To address these challenges, we introduce a modality that is inherently invariant to illumination and provides a consistent supervisory signal during training. Thermal imaging naturally exhibits this property. Unlike RGB cameras that capture reflected light, thermal sensors measure emitted radiation, offering reliable signals under low-light or even no-light conditions. Consequently, thermal data preserves stable contours and physically grounded spatial distributions correlated with material and temperature variations. We hypothesize that leveraging this illumination-invariant modality as a stable anchor can jointly guide RGB enhancement and 3D reconstruction toward a unified and consistent solution [11].

Based on this insight,We propose **DTGS**, a novel framework that tightly couples Retinex-inspired illumination enhancement [13] with end-to-end thermal-guided 3D Gaussian Splatting. Unlike prior pipelines that treat Retinex-style enhancement as a fixed pre-processing step, DTGS adopts a *joint* optimization strategy: illumination enhancement, 3D reconstruction, and thermal-specific consistency are optimized together. By closely coupling these components, DTGS can recover both structural and radiometric information from severely under-exposed inputs and maintain inter-view consistency even under the most challenging conditions.

The core innovation of DTGS lies in its integration of a *thermal supervisory branch* that jointly guides and stabilizes both Retinex-based enhancement and 3D Gaussian Splatting (3DGS) reconstruction. Within the cyclic enhancement–reconstruction process, the thermal modality provides illumination-invariant structural cues that constrain both reflectance recovery and geometric optimization. Meanwhile, an *adaptive weight scheduling* strategy dynamically balances the contributions of enhancement, geometric, and thermal losses, ensuring that illumination correction and 3D modeling evolve cooperatively rather than competitively. This unified design enables DTGS to achieve stable color restoration, robust radiometric consistency, and accurate geometry under extreme low-light conditions.

To rigorously evaluate DTGS, we construct a multi-view low-light thermal 3D reconstruction dataset. While existing datasets have explored NVS from raw noisy images [5, 16], a dedicated benchmark for thermal-guided 3D reconstruction under such adverse conditions is still lacking. Our dataset spans extremely adverse degradations, including ultra-low dynamic range, severe noise, and cross-view radiometric inconsistency, providing a realistic and challeng-

ing benchmark. Qualitative comparisons show that DTGS surpasses representative enhancement-plus-reconstruction baselines, achieving superior geometric fidelity and radiometric stability. These findings suggest that DTGS sets a new bar for robust 3D reconstruction in extremely low-light environments.

Our contributions are as follows:

- We propose a framework that tightly couples 3DGS with Retinex-based enhancement, trained under the joint supervision of a thermal modality, tailored for extreme low-light environments.
- We embed illumination-reflectance decomposition directly into the 3DGS loop, enabling physically-interpretable enhancement that co-evolves with geometry under thermal supervision.
- We propose a dynamic ground-truth update strategy that effectively breaks the enhancement-geometry deadlock by allowing the 3D model to train on progressively cleaner, internally-generated targets.
- We construct and release a challenging multi-view dataset, providing a realistic benchmark for evaluating illumination-aware and thermally-guided 3D reconstruction under adverse lighting conditions.

2. Related Work

2.1. Low-Light Enhancement.

Recent extensions of 3D Gaussian Splatting (3DGS), such as LITA-GS [34] and Luminance-GS [6], improve cross-view consistency via illumination-robust priors and view-adaptive tone curves, yet remain limited under extreme low light where SNR collapses and radiometric cues degrade. Gaussian-DK [31] incorporates exposure-compensated supervision for dark scenes but still suffers from color and geometric instability, highlighting the difficulty of jointly preserving photometry and structure. Retinex theory [13], including MSR/MSRCR, and its deep or curve-based variants such as RetinexNet [26], Zero-DCE [9], and ISP/Transformer approaches [4, 25], decompose images into reflectance and illumination. This decomposition is well suited for multi-view 3DGS: it separates exposure-variant illumination from exposure-invariant reflectance, aligns naturally with Gaussian color/opacity optimization, and allows priors like smooth illumination and sparse reflectance—together with auxiliary multi-view constraints (e.g., thermal histograms/gradients)—to prevent hallucination and maintain consistency in extreme low-light conditions.

2.2. Thermal Modality and Its Role in 3D Modeling

Thermal imaging complements RGB for 3D reconstruction by capturing emitted radiation, making it illumination-invariant and robust in low/no light [23, 27]. Early work

adapted SfM/SLAM to thermal inputs for point clouds and surfaces [7, 29], demonstrating geometric feasibility but limited texture fidelity due to lower resolution and higher noise.

Neural scene representations deepen these advances. NeRF variants fuse RGB/thermal streams or use thermal cues to guide geometry for robust novel-view synthesis [11, 28]. Gaussian splatting has likewise been extended with thermal data for real-time, high-quality rendering [2]. Cross-spectral formulations jointly model radiance and geometry with automatic pose calibration [19]; Thermal Gaussian [2], HyperGS [22], and SpectralGaussians [20] further improve radiometric consistency and geometric fidelity. Building on these, our method treats thermal signals as supervision within a 3DGS framework: thermal edges anchor Gaussian centers/covariances, histogram–gradient statistics enable per-view radiometric calibration, and high-SNR thermal gradients drive adaptive loss reweighting—stabilizing Retinex-style enhancement and reconstruction to yield color-consistent, geometrically faithful results under extreme low light.

3. Method

We propose DTGS, a 3D reconstruction framework tailored for low-light thermal imaging. Instead of relying on pixel-level enhancement, DTGS employs a Retinex-based decomposition that aligns with 3D Gaussian Splatting for joint optimization of reflectance, illumination, and geometry. To ensure consistency, we introduce a thermal consistency loss and a cyclic ground-truth refresh strategy, mitigating the "enhancement–geometry deadlock" and improving both radiometric and geometric fidelity. Thermal imaging further provides illumination-invariant structural cues, stabilizing the enhancement–reconstruction process and enabling sharper, more consistent results in extreme low-light scenes.

3.1. Cyclic Enhancement-Supervision Joint Optimization

The key innovation of our approach is the **cyclic enhancement-supervision mechanism**, which dynamically integrates image enhancement with 3D reconstruction in a mutually reinforcing loop. This mechanism targets a core challenge in low-light 3D reconstruction: a vicious cycle. Poor image quality hinders accurate geometry learning. In turn, inaccurate geometry limits the effectiveness of image enhancement. Our cyclic mechanism breaks this deadlock by allowing the ground truth to evolve progressively during training, as the enhancement network and 3D model co-adapt.

High-quality 3D reconstruction requires well-enhanced images, but effective image enhancement relies on accurate 3D geometry. To solve this, we introduce a novel **evolving supervision strategy** where the ground truth $GT^{(t)}$ is updated at each training iteration based on the output of the enhancement network.

3.1.1. Initialization and Cyclic Update Mechanism

At the start of training, the raw low-light images are used as the initial ground truth:

$$GT^{(0)} = I_{\text{low}}. \tag{1}$$

This serves as the starting point for the 3D Gaussian Splatting model. At each iteration t, the enhancement network $\mathcal{R}$ (Retinex) generates an enhanced image:

$$I_{\text{enh}}^{(t)} = \mathcal{R}(I_{\text{low}}). \tag{2}$$

The ground truth is then updated by blending the previous ground truth and the current enhanced output:

$$GT^{(t)} = (1 - \alpha^{(t)}) \cdot GT^{(t-1)} + \alpha^{(t)} \cdot I_{\text{enh}}^{(t)}, \tag{3}$$

where the blending factor $\alpha^{(t)}$ increases linearly to ensure smooth transition:

$$\alpha^{(t)} = \min\left(1.0, \frac{t}{T_{\text{transition}}}\right), \quad T_{\text{transition}} = 8{,}000. \tag{4}$$

the enhancement network typically converges around 5k–10k iterations independent of total training steps, a fixed threshold of 8,000 triggers the raw→enhanced supervision handover right after convergence—avoiding premature switches in short runs and unnecessary delays in long runs—yielding robust, efficient performance across scenes and budgets.

This multi-stage update ensures:

Early stage: Supervision relies primarily on raw input to avoid unstable enhancement artifacts. **Middle stage:** A balanced combination of raw and enhanced images stabilizes joint optimization. **Late stage:** Enhanced and thermally corrected images dominate, improving final radiometric and geometric quality.

3.1.2. Thermal & Geometric Losses with Adaptive Weighting

We define the total loss for the 3D Gaussian Splatting model as a combination of terms that balance image quality, structural consistency, and thermal fidelity.

To ensure radiometric coherence between the enhanced images and 3D reconstructions, we introduce a **thermal consistency loss** that preserves both the enhancement quality and the inherent thermal characteristics of the original imagery. Unlike conventional reconstruction losses that only measure pixel-wise similarity, our thermal consistency loss enforces a dual constraint that balances enhancement fidelity with thermal integrity:
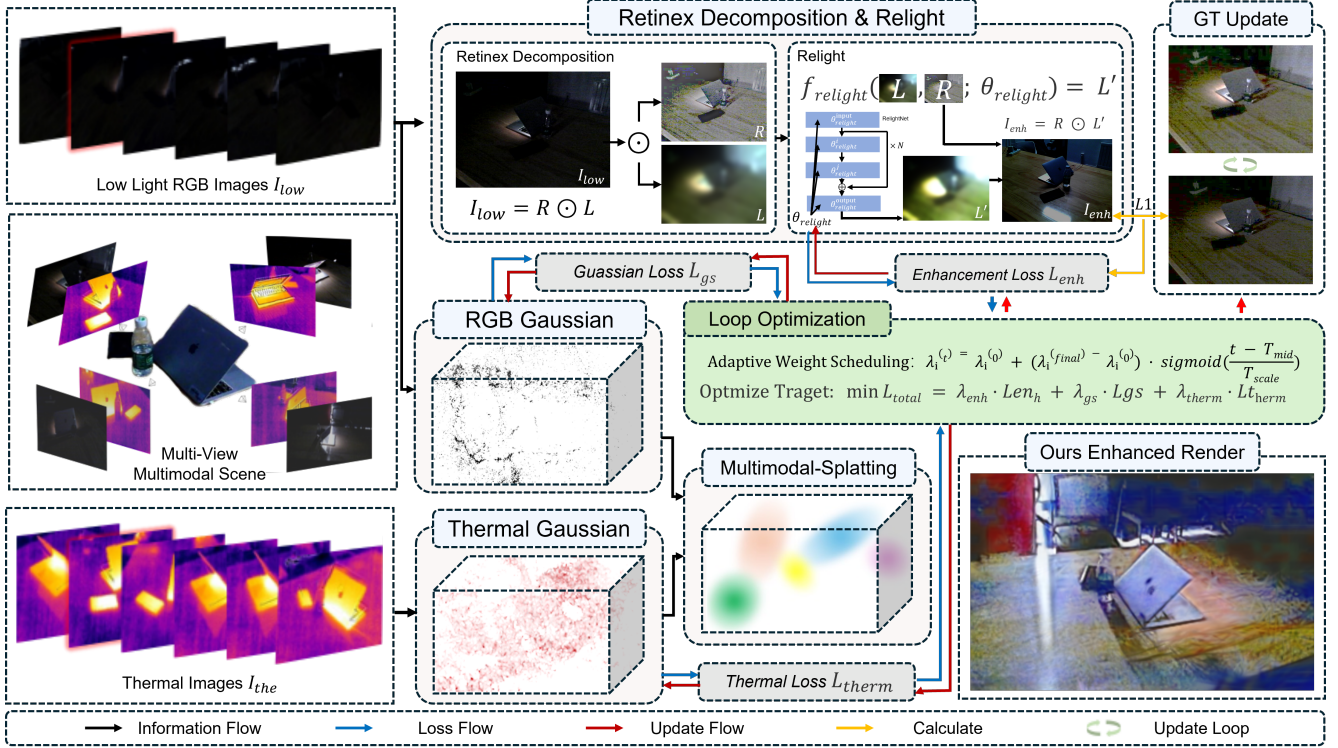
Figure 2. Overview of the DTGS Framework. Given multi-view low-light and thermal images of a scene, our method performs joint enhancement and 3D reconstruction through cyclic optimization. The thermal modality guides Retinex-based enhancement of low-light images, producing I_{enh} that progressively updates the ground truth $GT^{(t)} = (1 - \alpha^{(t)}) \cdot GT^{(t-1)} + \alpha^{(t)} \cdot I_{\text{enh}}^{(t)}$ for 3D Gaussian Splatting training. The total loss $\mathcal{L}_{\text{total}} = \lambda_{\text{enh}} \cdot \mathcal{L}_{\text{enh}} + \lambda_{\text{gs}} \cdot \mathcal{L}_{\text{gs}} + \lambda_{\text{therm}} \cdot \mathcal{L}_{\text{therm}}$ is optimized with adaptive four-stage weight scheduling. Thermal consistency loss ensures cross-modal coherence, enabling bidirectional gradient flow between enhancement and reconstruction modules.

$$\mathcal{L}_{\text{therm}} = \gamma \left\| \Phi_{\text{rgb}}(I_{\text{enh}}) - \Phi_{\text{rgb}}(I_{\text{rendered}}) \right\|_1$$
$$+ (1 - \gamma) \left\| \Phi_{\text{rgb} \leftrightarrow \text{therm}}(I_{\text{enh}}) - \Phi_{\text{therm}}(I_{\text{low}}^{\text{therm}}) \right\|_1 , \quad (6)$$

Here $I_{\text{low}}^{\text{therm}}$ denotes the original thermal infrared image (distinct from the low-light RGB I_{low}). We use lightweight cross-modal alignment to place both modalities in a shared metric space before the ℓ_1: Φ_{rgb} maps RGB to per-image luminance with min–max normalization, Φ_{therm} min–max normalizes the thermal image to $[0, 1]$ (or temperature range if calibrated), and $\Phi_{\text{rgb} \leftrightarrow \text{therm}}$ is the same luminance+normalization applied to I_{enh} so it matches the thermal range. This makes the two terms dimensionally consistent without changing network outputs. We set $\gamma = 0.1$.

The first term $\gamma \left\| \Phi_{\text{rgb}}(I_{\text{enh}}) - \Phi_{\text{rgb}}(I_{\text{rendered}}) \right\|_1$ enforces **enhancement-reconstruction consistency**, ensuring that the 3D Gaussian Splatting renderer produces outputs that align with the enhanced images. This cross-modal consistency is critical for preventing the "enhancement–geometry deadlock," where improved image quality leads to geometric artifacts, or vice versa. By maintaining this alignment, we ensure that geometric learning benefits from enhanced

visual features while remaining stable.

The second term provides **thermal preservation regularization**, preventing the enhancement network from over-adjusting the original thermal distribution. This constraint is particularly important for thermal imaging, where radiometric values carry physical meaning (e.g., temperature information). By penalizing excessive deviation from the raw input, we preserve the semantic content of thermal data while allowing necessary brightness and contrast improvements.

To ensure stable convergence and balanced optimization across multiple objectives, we formulate the **total training objective** that integrates three major components with **adaptive weighting**:

$$\mathcal{L}_{\text{total}}^{(t)} = \lambda_{\text{enh}}^{(t)} \, \mathcal{L}_{\text{enh}}(I_{\text{enh}}, I_{\text{low}})$$
$$+ \lambda_{\text{gs}}^{(t)} \, \mathcal{L}_{\text{gs}}(I_{\text{rendered}}, GT^{(t)}) \quad (7)$$
$$+ \lambda_{\text{therm}}^{(t)} \, \mathcal{L}_{\text{therm}},$$

where $GT^{(t)}$ evolves according to Eq. (3) with the linear blending schedule in Eq. (4), and the adaptive weights $\{\lambda_{\text{enh}}^{(t)}, \lambda_{\text{gs}}^{(t)}, \lambda_{\text{therm}}^{(t)}\}$ are dynamically adjusted throughout

training.

This scheduling allows us to shift focus from enhancing images early in training to refining geometry and thermal consistency later in the process.

3.1.3. Convergence and Benefits of the Cyclic Mechanism

The proposed cyclic training paradigm guarantees convergence by leveraging three core properties. Monotonicity: As the reliability weight $\alpha^{(t)}$ increases, the supervision progressively emphasizes higher-quality ground truth, ensuring a monotonic improvement of reconstruction fidelity. Stability: The gradual update strategy mitigates abrupt optimization shifts, thereby maintaining training stability. Adaptivity: The supervision dynamically aligns with the model's evolving prediction capability, facilitating self-consistent refinement.

Overall, this cyclic mechanism fosters mutual enhancement between image restoration and 3D geometry estimation, yielding reconstructions that are not only geometrically precise but also thermally coherent, even in complex low-illumination scenarios.

3.2. Retinex-Driven Joint Optimization with 3D Gaussian Splatting

To achieve illumination-aware 3D reconstruction under extremely low-light conditions, we formulate DTGS as a unified optimization framework that couples Retinex-based enhancement, thermal supervision, and 3D Gaussian Splatting (3DGS). Unlike conventional pipelines that apply enhancement as an independent preprocessing step, DTGS embeds Retinex decomposition directly inside the reconstruction loop, allowing illumination correction and geometric modeling to evolve jointly under thermal constraints.

3.2.1. Retinex formulation within 3DGS.

Given a low-light input image $I_{\text{low}} \in \mathbb{R}^{H \times W \times 3}$, DTGS decomposes it into reflectance R and illumination L following the Retinex principle:

$$I_{\text{low}} = R \odot L, \qquad I_{\text{enh}} = R \odot L'.$$

"L' denotes the illumination after correction.

To ensure smooth transition during training, we employ a **four-stage piecewise linear scheduling** mechanism for loss weights:

$$\lambda_k^{(t)} = \begin{cases} \lambda_k^{(0)}, & \text{if } t/T < 0.2 \\ \lambda_k^{(0)} + (\lambda_k^{(\text{final})} - \lambda_k^{(0)}) \cdot \frac{t/T - 0.2}{0.2}, & \text{if } 0.2 \leq t/T < 0.4 \\ \lambda_k^{(\text{final})}, & \text{if } 0.4 \leq t/T < 0.7 \\ \text{fine-tuned}, & \text{if } t/T \geq 0.7 \end{cases}$$

$$(8)$$

where T is the total number of iterations, and $k \in \{\text{enh}, \text{gs}, \text{therm}\}$. This four-stage design allows the model

to: (1) focus on enhancement initialization, (2) smoothly balance all objectives, (3) stabilize 3D reconstruction, and (4) fine-tune radiometric consistency. The weights satisfy the normalization constraint:

$$\lambda_{\text{enh}}^{(t)} + \lambda_{\text{gs}}^{(t)} + \lambda_{\text{therm}}^{(t)} = 1, \quad \lambda_{\text{gs}}^{(t)} \geq 0.1. \tag{9}$$

and the enhanced image is reconstructed as $I_{\text{enh}} = R \odot L'$. This decomposition explicitly separates exposure-variant illumination from exposure-invariant reflectance, producing physically interpretable enhancement that aligns with 3DGS appearance parameters (color and opacity). We initialize the task-level weight of the thermal branch at $\lambda_{\text{therm}}^{(0)} = 0.1$ and schedule it via Eq. (9) under the normalization constraint in Eq. (10).

When the thermal branch is enabled, task weights are always renormalized to satisfy Eq. (9). For the implementation detail "Retinex/3DGS set to $0.1/0.9$ with an additional thermal-specific 0.2", we use normalized weights $(\lambda_{\text{enh}}, \lambda_{\text{gs}}, \lambda_{\text{therm}}) = \frac{(0.1, 0.9, 0.2)}{0.1 + 0.9 + 0.2} \approx (0.083, 0.750, 0.167)$ to ensure $\lambda_{\text{enh}} + \lambda_{\text{gs}} + \lambda_{\text{therm}} = 1$.

3.2.2. Coupling with 3D Gaussian Splatting.

The enhanced image I_{enh} serves as supervision for the 3DGS renderer, which represents the scene as a set of N Gaussians $\mathcal{G} = \{(\mu_i, \Sigma_i, \alpha_i, c_i)\}_{i=1}^N$. Rendering is performed via front-to-back alpha compositing:

$$C(r) = \sum_{i=1}^{N(r)} c_i \alpha_i \prod_{j < i} (1 - \alpha_j), \tag{10}$$

where $C(r)$ is the color of ray r intersecting the Gaussian set. Loss gradients from 3D reconstruction propagate through the Retinex branch, forcing it to favor decompositions that yield geometry-consistent enhancement and stable cross-view illumination. This closed-loop coupling avoids the "enhancement–geometry deadlock" common in two-stage systems.

4. RGBT-LOW Dataset

To rigorously assess 3D reconstruction in low-light thermal conditions, we introduce the **RGBT-LOW** dataset. Existing benchmarks—such as those used in Thermal Gaussian—capture RGB images in well-lit environments, failing to represent severe real-world degradation. Datasets collected in dim settings often lack an additional modality like LOM [5], limiting their utility. RGBT-LOW addresses this gap by providing RGB scenes with minimal visible detail, forcing reconstruction methods to depend primarily on thermal cues.

The **RGBT-LOW** dataset consists of 20 real-world scenes, with a total of 6000 images, including various objects such as computer chairs, bottles, bags, and books.
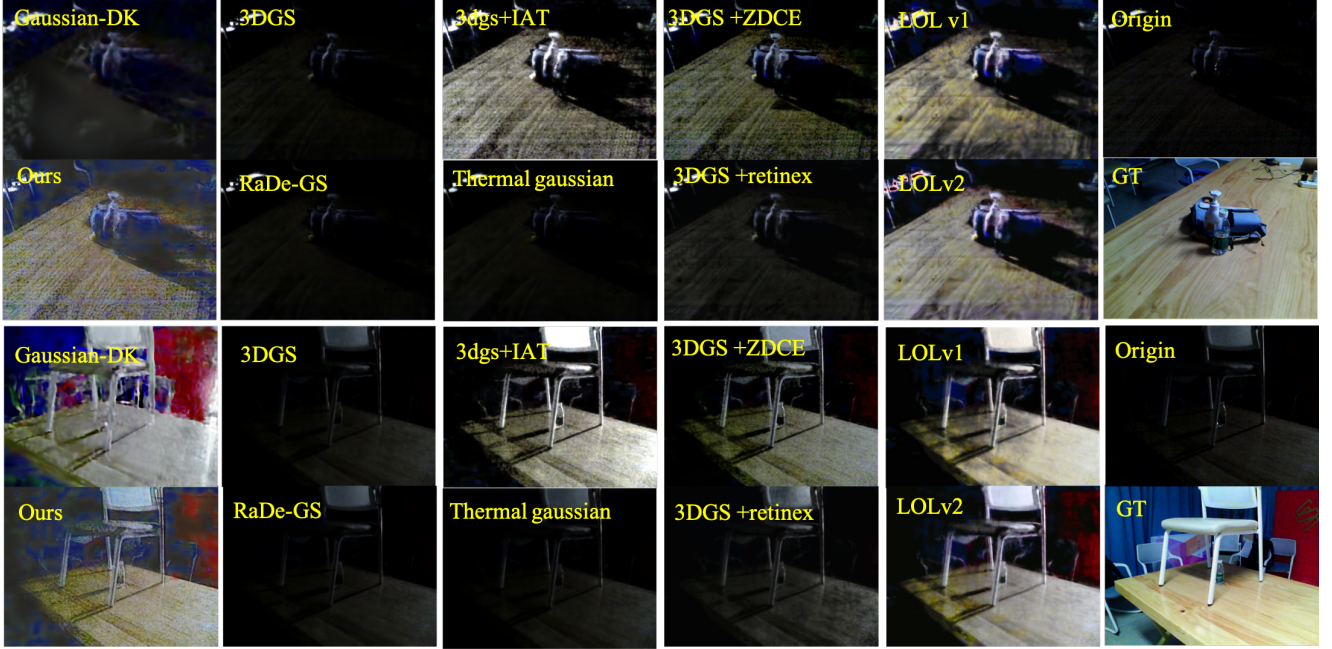
Figure 3. Qualitative comparison on the RGBT-LOW dataset. Our method produces more consistent color restoration and structural fidelity across views. Compared with other enhancement-based approaches, DTGS effectively preserves object integrity and avoids enhancement tearing, geometric distortion, or color shifts commonly seen in partially enhanced regions.include gaussian-DK [31],3dgs,retinexformer, RaDe-GS [32],Thermal Gaussian

This amounts to a total of $75\times2\times20\times2$ synchronized RGB (dark light corresponding to bright light GT) and thermal image pairs. For each viewpoint, we generate low-light and normal-light images, while keeping other camera settings unchanged. We capture multi-view images by moving and rotating the camera mount. The image acquisition resolution is 680×480. Each scene contains approximately $75\times2\times2$ images captured using **FLIR** cameras. Specifically, all images were captured using the **FLIR E6** camera model, which can simultaneously capture both RGB and thermal images.

To ensure fair comparisons, we use indoor lighting supplementation to collect the GT data and employ the same method for training, segmentation, and testing. We provide the original images captured by the thermal camera, along with RGB images, thermal images, MSX images, and camera pose data. Our dataset ensures consistency in thermal measurements across different viewpoints and covers a wide range of scenes. For more detailed information about our dataset, please refer to the supplementary materials.

5. Experiment

We evaluate all methods on the proposed RGBT-LOW dataset under consistent settings.

5.1. Implementation Details

All experiments are conducted on NVIDIA RTX 3090 GPUs. The model is trained for 30,000 iterations with an image resolution of 680×480. We use the Adam optimizer with an initial learning rate of 1×10^{-3} and apply cosine decay every 5000 iterations. The 3D Gaussian positions are initialized using COLMAP for camera pose estimation. In our joint training framework, the loss weights for RetinexNet and 3D Gaussian Structures (3DGS) are set to 0.1 and 0.9, respectively, along with a thermal-specific loss weight of 0.2 to address the unique characteristics of thermal images.

A cyclic ground truth (GT) update mechanism is employed, with a transition period of 8000 iterations. The overall loss function is a weighted combination of several components: L1 loss (0.7), SSIM loss (0.2), Corner loss (0.1), and consistency loss (0.1). In addition, adaptive weight scheduling is applied to RetinexNet. During early training (0–30%), the weight is set to 0.3, which de-

Table 1. Quantitative comparison of image quality and perceptual metrics, including SSIM,PSNR, and LPIPS,across four object classes and the overall mean. Section (A) reports 3D/thermal-only baselines, (B) presents pipelines combining low-light image enhancement with 3D Gaussian Splatting (3DGS), and (C) shows our proposed **DTGS** method. For fair comparison with recent low-light enhancement approaches, we additionally evaluate **Retinexformer** models like $LOLv1$ and $LOLv2real$ released in 2024–2025 (ECCV 2024 version and updates announced for the NTIRE 2025 Low-Light Image Enhancement Challenge.).The mean in this table is the average of multiple datasets, not just the datasets listed in the four columns below.*A detailed analysis of the selected baselines for comparison is provided in the Supplementary Material.*

Method	Books			Bottle			Chair			Computer			**mean**		
	SSIM ↑	PSNR ↑	LPIPS ↓	SSIM ↑	PSNR ↑	LPIPS ↓	SSIM ↑	PSNR ↑	LPIPS ↓	SSIM ↑	PSNR ↑	LPIPS ↓	SSIM ↑	PSNR ↑	LPIPS ↓
(A) 3D / Thermal-only baselines															
3DGS	0.2470	7.61	0.6844	0.0914	6.91	0.6793	0.1725	7.12	0.6802	0.1386	7.04	0.6837	0.1516	7.09	0.6865
RaDe-GS	0.2518	7.63	0.6891	0.0893	6.89	0.6813	0.1736	7.13	0.6802	0.1402	7.05	0.6846	0.1526	7.09	0.6882
Thermal Gaussian	0.2518	7.63	0.6891	0.1212	7.12	0.6811	0.1744	7.22	0.6721	0.0516	6.36	0.6932	0.1416	7.04	0.6877
Gaussian-DK	0.1886	7.63	0.6869	0.0755	6.96	0.6993	0.1271	7.15	0.6825	0.1043	7.03	0.7063	0.1134	7.09	0.7006
(B) Pre-enhance $\rightarrow$ 3DGS															
3DGS + IAT	0.2875	9.13	0.6623	0.1669	8.45	0.6875	0.2089	8.19	0.6790	0.2298	8.73	0.6791	0.2139	8.58	0.6811
3DGS + Zero-DCE	0.3010	10.18	0.6573	0.1381	7.85	0.6786	0.2045	8.74	0.6688	0.2120	8.70	0.6628	0.2034	8.69	0.6697
Zero-DCE + 3DGS	0.2920	9.30	0.6532	0.1381	7.92	0.6786	0.2055	8.92	0.6642	0.2298	8.92	0.6791	0.2053	8.68	0.6674
3DGS + Retinex	0.2790	8.41	0.6844	0.1631	7.40	0.6866	0.2299	7.71	0.6846	0.1978	7.59	0.6905	0.2076	7.67	0.6914
3DGS +LOL_v1	0.3444	9.32	0.6636	0.2101	8.23	0.6865	0.2668	8.42	0.6792	0.2869	8.74	0.6845	0.2776	8.68	0.6789
3DGS +LOL_v2_real	0.3498	9.66	0.6697	0.2552	8.89	0.6887	0.2774	8.73	0.6752	0.3061	9.26	0.6857	0.2971	9.13	0.6798
(C) Ours															
DTGS	**0.3251**	**11.07**	**0.6344**	**0.4842**	**11.17**	**0.6771**	**0.2689**	**10.64**	**0.6399**	**0.3445**	**11.70**	**0.6490**	**0.3520**	**11.17**	**0.6520**

creases to 0.2 in the mid-training phase (30–70%) and further reduces to 0.1 in the later stages (70–100%), ensuring a smooth transition from image enhancement to geometric refinement.

5.2. Comprehensive Evaluation

We conduct a comprehensive evaluation of novel view synthesis under low-light thermal conditions using the RGBT-LOW dataset. Initially, we evaluate the original 3DGS and RaDe-GS Thermal Gaussian on their ability to synthesize new views. Following this, to assess the effectiveness of our proposed end-to-end method in enhancing dark-light scenarios and preserving color consistency, we incorporate Zero-DCE and IAT for the original 3DGS and Gaussian-DK, which is another end-to-end network designed for enhancing dark-light thermal data.All enhancement models (IAT and Zero-DCE) are applied with their default pretrained weights and configurations.

5.2.1. Experimental Analysis

We conduct a comprehensive evaluation of the proposed **DTGS** framework under low-light conditions, comparing it against a range of state-of-the-art 3D Gaussian Splatting (3DGS) and low-light enhancement methods. Quantitative results are summarized in **Table 1**.

(A) 3D / Thermal-only Baselines. Traditional Gaussian-based rendering methods, including **3DGS**, **RaDe-GS**, **Thermal Gaussian**, and **Gaussian-DK** ("Gaussian in the Dark"), demonstrate highly limited capability in extremely dark scenes, with average SSIM around **0.15**, PSNR around **7 dB**, and LPIPS exceeding **0.68**. Notably, **Gaussian-DK** exhibits severe color inconsistency, geometric instability, and blurred rendering when applied to end-to-end low-light enhancement, indicating that the absence of an explicit illumination adaptation mechanism prevents these models from maintaining consistent geometry and texture fidelity under dark illumination. Even when thermal cues are available, conventional Gaussian rendering fails to preserve both *textural detail* and *lighting consistency* simultaneously.

(B) Pre-enhancement $\rightarrow$ 3DGS. To mitigate visibility degradation in low-light scenarios, we further evaluate combinations of representative enhancement models (**IAT**, **Zero-DCE**, **Retinex**, and **LOL** variants) as preprocessing modules before 3DGS reconstruction. Hybrid pipelines such as **3DGS+IAT** and **3DGS+Zero-DCE** show moderate gains in certain scenes (e.g., *books*, *computer*), with PSNR improving to **8.5–10 dB** and SSIM rising to **0.20–0.30**. However, their performance varies substantially across object categories. In complex or severely dark environments, they often cause *texture misalignment*, *geometric drift*, and *blurry point clouds*. For instance, both IAT and Zero-DCE fail to recover the red background in the *chair* scene, while **Gaussian-DK** produces local overexposure. Although recent transformer-based **Retinex** models (e.g., **LOL_v1**, **LOL_v2_real**) achieve improved brightness restoration and contrast balance (PSNR $\approx$ 8.7–9.1 dB, SSIM $\approx$ 0.27–0.30), they remain prone to *noise amplification* and *inconsistent color enhancement*, revealing unstable generalization in complex illumination.

(C) Ours: DTGS In contrast, our proposed DTGS achieves the best results across all datasets and metrics, with an average SSIM = 0.3520, PSNR = 11.17 dB, and LPIPS = 0.6520. Notably, relative to the strongest competing pipelines, DTGS improves the average SSIM by **+18.5%**, the average PSNR by **+22.3%**, and reduces LPIPS by **-2.3%**. On challenging scenes such as *books* and *bottle*, DTGS yields PSNR gains of +0.89 dB and +2.28 dB; for *chair* and *computer*, the gains are +1.72 dB and +2.44 dB, respectively. These consistent improvements highlight DTGS's capability to reconstruct fine geometry while maintaining thermal–visual coherence under extremely low-light conditions.

This remarkable performance stems from DTGS's cyclic enhancement–supervision mechanism, which enforces bidirectional consistency between the visual and thermal domains: (1) the enhancement branch leverages thermal cues to guide illumination restoration and noise suppression in the visual branch; (2) the supervision branch projects the enhanced visual features back to the thermal space to regularize geometry and structure; and (3) cyclic consistency losses align both domains at low-level (illumination/edges) and high-frequency (textures) scales. This dual-domain supervision enables effective recovery of low-level illumination cues and high-frequency detail while preserving heat-aware structural integrity.

Qualitative Evaluation. As illustrated in **Figure 3**, baseline methods such as **3DGS** and **RaDe-GS** produce blurred reconstructions with illumination artifacts, while enhancement-based variants such as **IAT** and **Retinexformer** yield partially recovered structures yet suffer from *uneven textures* and *color inconsistency*. In contrast, **DTGS** delivers *sharper edges*, *uniform brightness*, and *high-fidelity detail reconstruction*—for instance, clearly restoring the background panel and bottle contours. These results demonstrate that DTGS sets a new performance benchmark for low-light multimodal 3D Gaussian rendering, ensuring both **geometric accuracy** and **perceptual consistency** across diverse lighting conditions.

5.3. Ablation Study

To investigate the contribution of each module to the low-light thermal 3D reconstruction task, we perform a systematic ablation study on the RGBT-LOW dataset. The results are summarized in Table 2. Notably, **Thermal Gaussian** performs the worst across all metrics (average SSIM of only 0.1416, PSNR of 7.04 dB), showing that relying solely on thermal modality in extremely dark scenarios leads to significant information loss. Introducing cyclic supervision (**Thermal w/o Cyclic**) results in improvements in stability and light consistency, with a 3 dB increase in PSNR on average, validating the importance of cyclic GT updates for

Table 2. Ablation Study

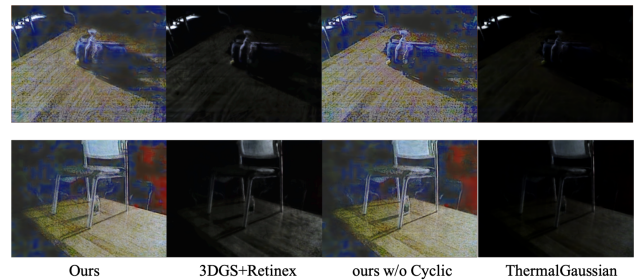| Class	Metric	Method			
		3DGS + Retinex	ours w/o Cyclic	Ours	Thermal Gaussian
bag	SSIM↑	0.2274	0.2431	**0.3368**	0.0716
	PSNR↑	7.67	10.37	**11.26**	6.65
	LPIPS↓	0.6897	0.6714	**0.6585**	0.7281
books	SSIM↑	0.2790	0.2104	**0.3251**	0.2518
	PSNR↑	8.41	8.90	**11.07**	7.63
	LPIPS↓	0.6844	0.6651	**0.6344**	0.6891
bottle	SSIM↑	0.1631	0.3207	**0.4842**	0.0893
	PSNR↑	7.40	10.95	**11.17**	6.89
	LPIPS↓	0.6866	**0.6756**	0.6771	0.6813
chair	SSIM↑	0.2299	0.1937	**0.2689**	0.1736
	PSNR↑	7.71	9.34	**10.64**	7.13
	LPIPS↓	0.6846	0.6617	**0.6399**	0.6802
computer	SSIM↑	0.1978	0.2283	**0.3445**	0.1402
	PSNR↑	7.59	10.00	**11.70**	7.05
	LPIPS↓	0.6905	0.6783	**0.6490**	0.6846
mean	SSIM↑	0.2076	0.2392	**0.3520**	0.1453
	PSNR↑	7.67	9.91	**11.17**	7.07
	LPIPS↓	0.6914	0.6704	**0.6520**	0.6927

Figure 4. Qualitative comparison of ablation variants.

joint optimization of geometry and thermal features.

Further, our method without the cyclic supervision (**Ours w/o Cyclic**) already demonstrates significant improvements over all baseline variants, confirming the effectiveness of the proposed thermal-aware cyclic enhancement mechanism. Compared with Thermal Gaussian and the pre-enhancement baseline (**3DGS + Retinex**), our DTGS model achieves the best performance, with an average SSIM of 0.3520 and PSNR of 11.17 dB. These results indicate that the integration of thermal guidance with cyclic ground-truth updates effectively mitigates degradation, enhances illumination stability, and maintains geometric consistency under extreme low-light conditions. Overall, the proposed framework establishes a robust foundation for illumination-aware 3D reconstruction.

6. Conclusion

We propose **DTGS**, the first thermal-supervised 3D Gaussian Splatting framework for extreme low light. DTGS unifies 3DGS reconstruction with a Retinex-based en-

hancer in a joint optimization, stabilized by a thermal branch that injects illumination-invariant structure to guide both reflectance recovery and geometry. A *Cyclic Enhancement–Supervision* scheme refreshes supervision during training so the two modules co-evolve.

References

[1] 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 *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12504–12513, 2023. 2

[2] Qian Chen, Shihao Shu, and Xiangzhi Bai. Thermal3dgs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis. In *European Conference on Computer Vision*, pages 253–269. Springer, 2024. 3

[3] Keith Richard Connor and Ian Reid. Novel view specification and synthesis. In *BMVC*, pages 1–10, 2002. 1

[4] Ziteng Cui, Kunchang Li, Lin Gu, Shenghan Su, Peng Gao, Zhengkai Jiang, Yu Qiao, and Tatsuya Harada. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. *arXiv preprint arXiv:2205.14871*, 2022. 2

[5] Ziteng Cui, Lin Gu, Xiao Sun, Xianzheng Ma, Yu Qiao, and Tatsuya Harada. Aleth-nerf: Illumination adaptive nerf with concealing field assumption, 2024. 2, 5

[6] Ziteng Cui, Xuangeng Chu, and Tatsuya Harada. Luminance-gs: Adapting 3d gaussian splatting to challenging lighting conditions with view-adaptive curve adjustment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26472–26482, 2025. 2

[7] Davide De Pazzi, Marco Pertile, and Sebastiano Chiodini. 3d radiometric mapping by means of lidar slam and thermal camera data fusion. *Sensors*, 22(21):8512, 2022. 3

[8] Noam Elata, Bahjat Kawar, Yaron Ostrovsky-Berman, Miriam Farber, and Ron Sokolovsky. Novel view synthesis with pixel-space diffusion models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26756–26766, 2025. 1

[9] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1780–1789, 2020. 1, 2

[10] Tewodros Habtegebrial, Kiran Varanasi, Christian Bailer, and Didier Stricker. Fast view synthesis with deep stereo vision. *arXiv preprint arXiv:1804.09690*, 2018. 1

[11] Mariam Hassan, Florent Forest, Olga Fink, and Malcolm Mielle. Thermonerf: Multimodal neural radiance fields for thermal novel view synthesis. *arXiv preprint arXiv:2403.12154*, 2(7), 2024. 2, 3

[12] 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, 2023. 1

[13] Edwin H Land and John J McCann. Lightness and retinex theory. *Journal of the Optical society of America*, 61(1):1–11, 1971. 2

[14] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE transactions on pattern analysis and machine intelligence*, 44(8):4225–4238, 2021. 1

[15] Zuoyue Li, Tianxing Fan, Zhenqiang Li, Zhaopeng Cui, and Yoichi Sato. Compnvs: Novel view synthesis with scene completion. *arXiv preprint arXiv:2207.11467*, 2022. 1

[16] Rongfeng Lu, Hangyu Chen, Zunjie Zhu, Yuhang Qin, Ming Lu, Le Zhang, Chenggang Yan, and Anke Xue. Thermal-gaussian: Thermal 3d gaussian splatting. *arXiv preprint arXiv:2409.07200*, 2024. 2

[17] 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. *Communications of the ACM*, 65(1):99–106, 2021. 1

[18] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16190–16199, 2022. 1

[19] Matteo Poggi, Pierluigi Zama Ramirez, Fabio Tosi, Samuele Salti, Stefano Mattoccia, and Luigi Di Stefano. Cross-spectral neural radiance fields. In *2022 International Conference on 3D Vision (3DV)*, pages 606–616. IEEE, 2022. 3

[20] Saptarshi Neil Sinha, Holger Graf, and Michael Weinmann. Spectralgaussians: Semantic, spectral 3d gaussian splatting for multi-spectral scene representation, visualization and analysis. *arXiv preprint arXiv:2408.06975*, 2024. 3

[21] Hao Tang, Hongyu Zhu, Huanjie Tao, and Chao Xie. An improved algorithm for low-light image enhancement based on retinexnet. *Applied Sciences*, 12(14):7268, 2022. 2

[22] Christopher Thirgood, Oscar Mendez, Erin Ling, Jon Storey, and Simon Hadfield. Hypergs: Hyperspectral 3d gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5970–5979, 2025. 3

[23] Stephen Vidas, Peyman Moghadam, and Michael Bosse. 3d thermal mapping of building interiors using an rgb-d and thermal camera. In *2013 IEEE international conference on robotics and automation*, pages 2311–2318. IEEE, 2013. 2

[24] Haoyuan Wang, Xiaogang Xu, Ke Xu, and Rynson WH Lau. Lighting up nerf via unsupervised decomposition and enhancement. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12632–12641, 2023. 1

[25] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI conference on artificial intelligence*, pages 2654–2662, 2023. 2

[26] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 2

[27] Martin Weinmann, Jens Leitloff, Ludwig Hoegner, Boris Jutzi, Uwe Stilla, and Stefan Hinz. Thermal 3d mapping

for object detection in dynamic scenes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2:53–60, 2014. 2

[28] Jiacong Xu, Mingqian Liao, Ram Prabhakar Kathirvel, and Vishal M Patel. Leveraging thermal modality to enhance reconstruction in low-light conditions. In *European Conference on Computer Vision*, pages 321–339. Springer, 2024. 3

[29] Masahiro Yamaguchi, Hideo Saito, and Shoji Yachida. Application of lsd-slam for visualization temperature in wide-area environment. In *International Conference on Computer Vision Theory and Applications*, pages 216–223. SciTePress, 2017. 3

[30] Mao-xiang Yang, Gui-jin Tang, Xiao-hua Liu, Li-qian Wang, Zi-guan Cui, and Su-huai Luo. Low-light image enhancement based on retinex theory and dual-tree complex wavelet transform. *Optoelectronics Letters*, 14(6):470–475, 2018. 2

[31] Sheng Ye, Zhen-Hui Dong, Yubin Hu, Yu-Hui Wen, and Yong-Jin Liu. Gaussian in the dark: Real-time view synthesis from inconsistent dark images using gaussian splatting. In *Computer Graphics Forum*, page e15213. Wiley Online Library, 2024. 2, 6

[32] Baowen Zhang, Chuan Fang, Rakesh Shrestha, Yixun Liang, Xiaoxiao Long, and Ping Tan. Rade-gs: Rasterizing depth in gaussian splatting. *arXiv preprint arXiv:2406.01467*, 2024. 6

[33] Minglu Zhang, Yan Zhang, Zhihong Jiang, Xiaoling Lv, and Ce Guo. Low-illumination image enhancement in the space environment based on the dc-wgan algorithm. *Sensors*, 21 (1):286, 2021. 1

[34] Han Zhou, Wei Dong, and Jun Chen. Lita-gs: Illumination-agnostic novel view synthesis via reference-free 3d gaussian splatting and physical priors. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21580–21589, 2025. 2