Physically Grounded Dual-Opacity Gaussian Splatting for Joint RGB-TIR Reconstruction

Jin Liu[1,2*], Dabin Leng[1,2*], Jiagang Chen[3], Haodong Li[1,2], Jiguang Li[4], Zhao Huang[5], Xiaoshuai Zhang[6], Qi Xu[7†], Zhiwen Zheng[1,2†], and Xingru Huang[1,2†]

[1] Hangzhou Dianzi University, Hangzhou, China
[2] Weidian Corporation, Beijing, China
[3] Wuhan University, Wuhan, China
[4] Northumbria University, Newcastle upon Tyne, UK
[5] University of Aberdeen, Aberdeen, UK
[6] Ocean University of China, Qingdao, China
[7] Westlake University, Hangzhou, China

Abstract. 3D Gaussian Splatting (3DGS) enables efficient novel-view synthesis, yet its extension to simultaneous RGB and thermal infrared (TIR) reconstruction is hindered by a fundamental spectral mismatch: RGB relies on reflected illumination while TIR captures emitted radiance. This disparity in frequency characteristics and physical sensing mechanisms causes gradient interference when naively sharing a unified representation. Common methods typically treat TIR as an auxiliary intensity channel, ignoring the physical radiative transfer governing thermal imaging. Critically, enforcing a single opacity field across modalities is physically invalid due to cross-spectral visibility disparities, constraining representational capacity and yielding thermally implausible reconstructions. To address these limitations, we propose Physically Grounded Dual-Opacity Gaussian Splatting, a framework that unifies RGB and TIR reconstruction under a shared geometric scaffold with modality-aware visibility modeling. We introduce Radiative Attribute Parameterization, explicitly modeling each Gaussian's thermal response through emissivity, temperature, and reflectance. Dual-Opacity Rendering enables spectrally aware visibility handling by assigning modality-specific opacities to shared location parameters, resolving cross-modal occlusion conflicts. Training jointly optimizes photometric reconstruction, physical priors, and RGB-guided geometric regularization to resolve parameter ambiguities. Experiments demonstrate competitive performance in both radiometric accuracy and photorealistic novel view synthesis, highlighting the potential of 3DGS for joint thermal field with high-fidelity RGB reconstruction.

Keywords: 3D Gaussian Splatting · RGB–TIR Reconstruction · Physics-grounded Rendering

* Equal Contribution.
† Corresponding Author.

1 Introduction

With the rapid expansion of digital twins and autonomous driving, all-weather and robust multi-modal 3D scene reconstruction has become a central requirement in computer vision [45]. Despite steady progress in RGB-based reconstruction in terms of geometric detail and texture fidelity, RGB observations alone are often constrained in extreme conditions, such as low illumination, strong camouflage, or smoke occlusion, where the scene's semantic cues and physical attributes are only partially observable [14,35]. Thermal infrared (TIR) imagery, by sensing radiance induced by surface temperature [10], provides complementary measurements that are largely independent of ambient lighting and expose thermal distributions beyond RGB spectrum [7]. Specifically, TIR sensors capture the long-wave infrared radiation emitted by objects, allowing for the detection of targets based on thermal emissivity differences even in total darkness. This modality is particularly effective at resolving thermal silhouettes and heat-conductive patterns that are invisible to standard optical sensors. Joint reconstruction from RGB and TIR therefore supports not only visually plausible 3D models but also physically grounded temperature attributes, which are highly relevant to critical applications including disaster response, nighttime surveillance, and industrial inspection [4, 13, 21].

The central challenge of RGB-TIR joint reconstruction arises from the intrinsic disparity between RGB and TIR in both imaging formation and spectral characteristics [27, 29]. Existing multi-modal 3D reconstruction methods, particularly extensions of Neural Radiance Fields (NeRF) [24] and 3D Gaussian Splatting (3DGS) [19, 41], are often constrained to improving rendering quality within a single modality, while leaving the modality specific formation mechanisms under-modeled. Consequently, thermal observations are frequently treated as an auxiliary single channel intensity signal, and optimized by directly reusing appearance models developed for the visible spectrum.

TIR imaging follows radiative physics (e.g., the Stefan-Boltzmann law), where the measured signal is jointly governed by surface emissivity, temperature, and reflected ambient radiation rather than by a purely illumination reflection model [34]. Treating TIR as a intensity mapping detaches the observation from its radiometric meaning [43]. In addition, RGB images typically exhibit rich high frequency textures, whereas TIR images often present low frequency, piecewise smooth thermal patterns, yielding mismatched gradient profiles at object boundaries and interior structures [15,17]. When a single geometric opacity is enforced across both modalities, the conflict between high and low frequency supervision can induce geometric artifacts during joint optimization, leaving a persistent trade off between true color detail and thermal fidelity [30].

To tackle these challenges, we propose RaViGS, a bi-modal 3D Gaussian Splatting framework for joint RGB-TIR reconstruction. RaViGS models both modalities in parallel on a unified geometric scaffold: the RGB branch captures view-dependent appearance, while the TIR branch factorizes thermal radiance into intrinsic attributes, e.g., emissivity, temperature, and reflectance, explicitly accounting for both emission and environmental reflection. To mitigate cross-spectral structural conflicts, we introduce a spectrally decoupled dual-opacity rendering mechanism. This scheme allows primitives to maintain independent visibility and transmittance for each modality while retaining joint supervision over shared geometric parameters. Finally, a joint optimization loss, incorporating bi-modal reconstruction, physical priors, and edge-guided smoothness, is formulated within a unified framework to preserve true-color details and ensure radiometrically consistent thermal recovery. Experiments demonstrate that the proposed method achieves consistent quality improvements across both modalities and ensures stable novel-view synthesis under complex illumination and thermal environments.

The contributions are summarized as follows:

1. We propose RaViGS, a joint RGB-TIR reconstruction framework that integrates radiative physical priors to resolve cross-spectral gradient conflicts and ensure radiometric consistency.
2. We devise Intrinsic Radiative Attribute Factorization and a dual-opacity mechanism to parameterize TIR formation via physical attributes while decoupling modality-specific visibility to suppress silhouette artifacts.
3. Comprehensive experiments on public datasets demonstrate that RaViGS achieves competitive performance in bi-modal novel-view synthesis across diverse challenging scenes.

2 Related Work

2.1 Novel View Synthesis

Transitioning from implicit volumetric formulations, parameterized by MLPs [20, 24] and augmented via conical integration or hash encoding [25], novel view synthesis has increasingly adopted explicit point primitive representations. Specifically, 3DGS [16] utilizes anisotropic primitives and differentiable rasterization , balancing rendering fidelity with computational throughput [1,2]. While subsequent extensions have systematically refined this architecture to address frequency stability, surface consistency, and high dimensional feature embedding [28, 44], these frameworks remain constrained to visible spectrum appearance regression. By leaving intrinsic physical attributes, such as temperature and material emissivity, under-modeled, existing paradigms are challenged by radiometric modalities [6,36,39]. Direct transfer of these color centric models to TIR observations thereby bypasses authentic heat-radiation mechanisms, fundamentally limiting the physical interpretability of the reconstructed scenes [26].

2.2 TIR Scene Reconstruction

TIR 3D reconstruction provides critical observability in visually degraded environments where visible spectrum sensors often fail [23]. Early methodologies adapted Structure-from-Motion (SfM) [32] and Multi-View Stereo (MVS) [31] pipelines by introducing robust cross spectral matching descriptors to handle the non-linear intensity variations between modalities [5]. Recently, neural rendering architectures, including implicit NeRF-based formulations [8,38] and explicit 3DGS frameworks [3, 21, 37], have facilitated high-fidelity TIR synthesis through differentiable rasterization. However, current paradigms predominantly regress coupled radiance intensity or pseudo-color maps. They lack an explicit, identifiable decomposition of underlying radiometric components [38]. Consequently, stably decoupling the physical temperature field from inherent material attributes under multi-view supervision remains an unresolved challenge [26].

2.3 Multi-modal Scene Reconstruction

The objective of joint multi-modal reconstruction is to leverage complementary signals to concurrently enhance geometric accuracy and cross-spectral representation [46]. Early RGB-TIR fusion methods primarily focused on 2D image space registration and texture mapping [12,18]. In the context of neural rendering, recent advancements have employed dual-stream encoders and cross modal attention mechanisms to aggregate features, thereby strengthening structural supervision and refining edge quality [46]. methods based on the 3DGS [16] paradigm demonstrate the feasibility of joint modeling within a unified spatial scaffold. However, while a shared geometric scaffold is effective for multi-modal alignment, existing methods suffer from tightly coupled visibility parameters. In regions with pronounced disparities in spectral transmission and occlusion (e.g., glass or camouflage), this coupling often induces boundary ghosting, local structural degradation, and over-smoothed thermal fields [38]. To bridge this gap, RaViGS introduces modality-decoupled visibility under a shared geometric framework and integrates infrared radiative factorization into a unified model.

3 The Proposed RaViGS

RaViGS is a physically-grounded framework for joint RGB-TIR reconstruction within a unified 3D Gaussian representation. As shown in Fig. 1, each primitive shares a common geometric scaffold while maintaining modality-specific appearance and visibility. Specifically, the RGB branch captures view-dependent color, whereas the TIR branch employs *Intrinsic Radiative Attribute Factorization* to parameterize thermal response based on radiative physics. To accommodate varying spectral transmittances, a *Dual-Opacity Rendering Scheme* accumulates visibility separately for each modality, confining cross-spectral conflicts to the radiative and visibility layers. The pipeline is optimized via a joint objective integrating bi-modal reconstruction, structural regularization, and physical priors. This parallel structure ensures shared geometric consistency while retaining high-fidelity visual details and radiometrically accurate thermal modeling.

3.1 Preliminaries of 3D Gaussian Splatting

3DGS [16] explicitly parameterizes a scene using a set of discrete anisotropic 3D Gaussian primitives. Each primitive is represented as a collection of geometric

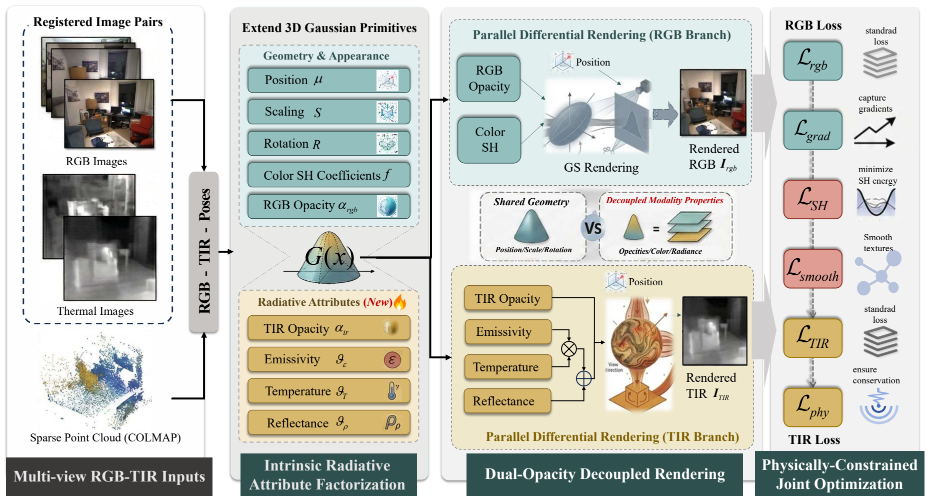

Fig. 1: Overview of the RaViGS method. RaViGS is a physically-grounded framework for joint RGB-TIR reconstruction within a unified 3D Gaussian representation. Each Gaussian primitive shares a common geometric scaffold while maintaining modality-specific appearance and visibility. The RGB branch captures view-dependent color, and the TIR branch employs Intrinsic Radiative Attribute Factorization to parameterize thermal response based on radiative physics. A Dual-Opacity Rendering Scheme separately accumulates visibility for each modality to confine cross-spectral conflicts. The entire pipeline is optimized via a joint objective integrating bi-modal reconstruction, structural regularization, and physical priors.

and appearance attributes:

$$\mathcal{G}_i = \{\mu_i, R_i, S_i, \alpha_i, f_i\}, \tag{1}$$

where $\mu_i \in \mathbb{R}^3$ denotes the primitive center, R_i and S_i are rotation and scaling matrices, and $\alpha_i \in [0, 1]$ is a scalar opacity parameter. Crucially, f_i represents the spherical harmonics (SH) coefficients used to encode view-dependent appearance. The spatial contribution of a primitive at a 3D location $x \in \mathbb{R}^3$ is described by an unnormalized Gaussian:

$$G(x) = \exp\left(-\tfrac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right), \tag{2}$$

where the covariance is given by $\Sigma = RSS^\top R^\top$.

During rendering, 3D Gaussians are projected onto the 2D image plane and accumulated via point based forward α-blending. For a pixel p, the rendered color $C(p)$ is computed as

$$C(p) = \sum_{i \in \mathcal{N}(p)} c_i(d)\, \alpha_i(p) \prod_{j=1}^{i-1} (1 - \alpha_j(p)), \tag{3}$$

where $\mathcal{N}(p)$ denotes the depth sorted set of Gaussians contributing to p, $c_i(d)$ is the SH evaluation under viewing direction d, and $\alpha_i(p)$ is the effective opacity after projection, coupled with the local point density. This formulation supports efficient modeling of visible spectrum appearance, while its parameterization and rendering assumptions do not directly encode TIR radiative formation.

3.2 Intrinsic Radiative Attribute Factorization

Unlike visible imaging, which is dominated by the reflection of external illumination, the signal recorded by a TIR sensor is not a direct grayscale mapping of surface appearance, but a superposition of self emitted thermal radiation and reflected ambient radiance. Consequently, in several multi-modal reconstruction pipelines, treating TIR intensity as an additional color channel for standard 3DGS optimization misaligns the representation with radiative physics, and tends to render the TIR branch under-determined, often accompanied by artifacts in novel view synthesis. To bridge this gap, we reformulate the Gaussian primitive representation by factorizing the TIR response into intrinsic physical attributes, shifting the learning objective from heuristic intensity regression to physics-grounded radiometric inference.

Following the Stefan-Boltzmann law, we represent each 3D Gaussian $\mathcal{G}_i$ using three learnable parameters with explicit physical interpretations: emissivity ϵ, normalized temperature T, and reflectance ρ. Although these attributes are intrinsic to the material, the effective radiance of complex scenes often exhibits directional dependency due to non-Lambertian emission on unpolished surfaces, micro-facet shadowing effects, and local thermal gradients. Consequently, we parameterize these attributes using Spherical Harmonics to capture such anisotropic radiative behavior. For a primitive i observed from direction d, the view-dependent attributes are defined as:

$$\epsilon_i(d) = \mathcal{A}\left(\Phi(\theta_i^{\epsilon}, d)\right), \quad T_i(d) = \mathcal{A}\left(\Phi(\theta_i^{T}, d)\right), \quad \rho_i(d) = \mathcal{A}\left(\Phi(\theta_i^{\rho}, d)\right), \tag{4}$$

where $\Phi(\cdot, d)$ denotes the SH evaluation, $\theta_i^{\{\cdot\}}$ are the learned coefficients, and $\mathcal{A}(\cdot)$ is a sigmoid activation ensuring the attributes remain within the physically plausible range of $[0, 1]$.

During rendering, the per-primitive TIR radiative intensity along d is modeled as the sum of an emission term and an ambient-reflection term:

$$E_i(d) = \mathrm{clip}\left(\epsilon_i(d) \cdot \sigma T_i(d)^4 + \rho_i(d) \cdot L_{\mathrm{env}}, [0, 1]\right), \tag{5}$$

where T_i^4 models the non-linear energy-temperature dependence, σ is a learnable scale factor absorbing the Stefan-Boltzmann constant and sensor gain, and L_{env} is a global parameter approximating omnidirectional ambient thermal radiation. The computed $E_i(d)$ is then integrated along the ray using the same volumetric compositing as the RGB branch.

With this explicit radiative factorization, Gaussian attributes are no longer optimized to memorize TIR grayscale values, but to parameterize material related radiative properties and thermodynamic state, supporting thermal texture reconstruction with improved physical consistency.

3.3 Dual-Opacity Decoupled Rendering

In multi-modal scenes, electromagnetic waves at different bands follow distinct spectral behaviors, under which occlusion and transmission can differ markedly between RGB and TIR observation. A common case is glass, which is largely transparent in RGB ($\alpha_{rgb} \to 0$) yet strongly absorbing and thus effectively opaque in long-wave TIR ($\alpha_{ir} \to 1$). When a single opacity parameter α is enforced across modalities, it is simultaneously supervised by inconsistent signals, which can induce conflicting gradients and manifest as geometric artifacts.

To accommodate these cross-spectral visibility misalignments while preserving a unified geometric backbone, a spectrally decoupled dual-opacity rendering strategy is adopted, where each Gaussian primitive maintains modality-specific opacity fields. Specifically, the i-th primitive is parameterized as

$$\mathcal{G}_i = \{\mu_i, R_i, S_i, f_i, \alpha_i^{\mathrm{rgb}}, \theta_i^{\epsilon}, \theta_i^{T}, \theta_i^{\rho}, \alpha_i^{\mathrm{ir}}\}, \tag{6}$$

where f_i denotes the SH coefficients for RGB appearance; $\theta_i^{\epsilon}, \theta_i^{T}, \theta_i^{\rho}$ denote the SH coefficients of the thermal physical attributes in Sec. 3.2; and $\alpha_i^{\mathrm{rgb}}, \alpha_i^{\mathrm{ir}} \in [0, 1]$ correspond to the opacities of RGB and TIR, respectively.

Given a pixel p and viewing direction d, let $\mathcal{N}(p)$ denote the depth sorted set of contributing Gaussians. Both branches share the same projection and depth ordering, yet accumulate transmittance with independent opacities via forward α-blending:

$$\hat{C}(p) = \sum_{i \in \mathcal{N}(p)} T_i^{\mathrm{rgb}}(p) \alpha_i^{\mathrm{rgb}}(p) c_i(d), \tag{7}$$

$$\hat{I}_{\mathrm{ir}}(p) = \sum_{i \in \mathcal{N}(p)} T_i^{\mathrm{ir}}(p) \alpha_i^{\mathrm{ir}}(p) E_i(d), \tag{8}$$

$$T_i^{m}(p) = \prod_{j<i} \big(1 - \alpha_j^{m}(p)\big), \quad m \in rgb, ir, \tag{9}$$

where $c_i(d)$ is the SH-evaluated RGB color, and $\Phi_{ir}(i, d)$ is the per-Gaussian thermal radiance term from Sec. 3.2. This parallel compositing routes modality specific residuals primarily to branch parameters, while geometric variables (position, rotation, scale) receive joint supervision from both modalities:

$$\frac{\partial \mathcal{L}}{\partial \theta_{geo}} = \lambda_{rgb} \frac{\partial \mathcal{L}_{rgb}}{\partial \theta_{geo}} + \lambda_{ir} \frac{\partial \mathcal{L}_{ir}}{\partial \theta_{geo}}. \tag{10}$$

Under this dual-gradient constraint, the underlying geometry is regularized by bi-modal evidence to ensure structural consistency, whereas spectral differences are absorbed by modality specific opacity and radiative parameters.

To prevent the training process from degenerating into two modality-specific Gaussian subsets, we couple the densification and pruning processes through the co-support criterion. Specifically, we maintain two learnable opacity fields, α_i^{rgb} and α_i^{ir}, for the RGB and TIR rendering paths respectively. During the k-th density update iteration, if a Gaussian primitive becomes no longer significant in either modality, it is pruned:

$$M^{(k)}_{\text{prune}}(i) = 1\Big[\min\big(\alpha_i^{rgb}, \alpha_i^{ir}\big) < \tau_\alpha\Big]. \tag{11}$$

where $1(\cdot)$ is the indicator function and τ_α is the opacity threshold. This min-gating strategy forces each retained Gaussian primitive to be jointly supported by both modalities, thereby suppressing the optimizer from allocating independent primitives that only explain single-modality observations (only RGB or only thermal infrared). This strategy produces a single shared geometric skeleton while allowing both modalities to express appearance differences through their respective independent opacity and physical parameter branches.

3.4 Physically-Constrained Joint Optimization

To resolve the ambiguities in radiative attribute factorization and cross-spectral geometric alignment, we formulate a multi-objective optimization. This joint objective orchestrates photometric reconstruction, structural regularization, and thermodynamic priors to effectively supervise the dual-opacity fields and the intrinsic physical parameters.

Let Ω denote the pixel set and $\mathcal{K}$ the Gaussian set. The rendered outputs are I_{rgb} and I_{ir} for the RGB and TIR branches, respectively.

$$\begin{aligned}\mathcal{L}^m_{rec} = \frac{1}{|\Omega|}\sum_{p\in\Omega}\Big[&(1-\lambda_{dssim})\|I_m(p) - I_m^{gt}(p)\|_1 \\ &+ \lambda_{dssim}\mathrm{DSSIM}\big(I_m(p), I_m^{gt}(p)\big)\Big], m \in \{rgb, ir\}\end{aligned} \tag{12}$$

To suppress high frequency noise in the RGB branch and ensure smooth novel view rendering, we introduce a gradient alignment loss $\mathcal{L}_{grad}$ and an SH-coefficient regularizer $\mathcal{L}_{sh}$:

$$\mathcal{L}_{\text{grad}} = \frac{1}{2|\Omega|}\sum_{p\in\Omega}\Big(\|\nabla_x I_{\text{rgb}}(p) - \nabla_x I^{\text{gt}}_{\text{rgb}}(p)\|_1 + \|\nabla_y I_{\text{rgb}}(p) - \nabla_y I^{\text{gt}}_{\text{rgb}}(p)\|_1\Big) \tag{13}$$

$$\mathcal{L}_{\text{sh}} = \frac{1}{|\mathcal{K}|}\sum_{i\in\mathcal{K}}\|f_{\text{rest},i}\|_2^2, \tag{14}$$

where $f_{rest,k}$ are the higher-order SH coefficients of the k-th Gaussian. $\mathcal{L}_{grad}$ constrains image edges, and $\mathcal{L}_{sh}$ limits high frequency component energy to reduce view specific overfitting.

TIR intensity is determined by temperature, emissivity, and reflectance. To ensure physically consistent parameterization, we introduce a physical prior loss $\mathcal{L}_{phy}$ based on Kirchhoff's law:

$$\mathcal{L}_{\text{phy}} = \frac{1}{|\mathcal{K}|}\sum_{i\in\mathcal{K}}\big(\mathcal{B}(\epsilon_i^0)^2 + \mathcal{B}(\rho_i^0)^2 + \mathrm{ReLU}(\epsilon_i^0 + \rho_i^0 - (1+\delta))^2\big) \tag{15}$$

where $\mathcal{B}(x) = \mathrm{ReLU}(-x) + \mathrm{ReLU}(x-1)$ enforces the bounding constraint for emissivity and reflectance. Here, ϵ_i^0 and ρ_i^0 represent the 0-th order components of the SH coefficients, enforcing the physical prior $\epsilon_i^0, \rho_i^0 \in [0,1]$ on the base material attributes. A tolerance $\delta = 0.05$ is applied for minor transmission and sensor nonlinearity.

Given the low resolution and sparse texture of TIR images, we use RGB gradients as geometric guidance through an edge aware smoothness loss $\mathcal{L}_{smooth}$:

$$\mathcal{L}_{smooth} = \frac{1}{|\Omega|} \sum_{p \in \Omega} \|\nabla I_{ir}(p)\|_1 \cdot \exp\left(-\beta \|\nabla I_{rgb}^{gt}(p)\|_1\right). \tag{16}$$

Here, strong RGB gradients act as a gating signal: the constraint is relaxed at RGB edges to preserve sharper TIR boundaries and enforced in flat regions to suppress thermal noise.

The total objective is given by:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rgb}} \left(\mathcal{L}_{\text{rec}}^{\text{rgb}} + \lambda_g \mathcal{L}_{\text{grad}} + \lambda_{\text{sh}} \mathcal{L}_{\text{sh}}\right) + \lambda_{\text{ir}} \left(\mathcal{L}_{\text{rec}}^{\text{ir}} + \lambda_s \mathcal{L}_{\text{smooth}} + \lambda_p \mathcal{L}_{\text{phy}}\right). \tag{17}$$

This joint optimization constrains the RGB and TIR branches under shared geometry and uses physical priors to reduce ambiguity in thermal parameter factorization.

4 Experiments

We conduct extensive experiments across various challenging indoor and outdoor scenes, comparing our approach against representative multi-modal novel view synthesis methods.

4.1 Experimental Settings

Dataset. ThermoNeRF [9] dataset is captured using an iPhone-mounted FLIR One Pro LT stereo camera, including multi-view RGB and TIR image sequences at both object and scene scales. The dataset contains 16 scenes spanning different temperature ranges, partitioned into 8 building-façade scenes and 8 additional daily scenes. For each scene, the device provides synchronized and strictly aligned RGB-TIR image pairs. For reference temperature, although the device specification reports a thermal accuracy of $\pm 3°$C within $-20°$C to $120°$C, the dataset authors empirically assess an effective thermal accuracy of about $\pm 0.14°$C. These temperature measurements are used as the primary reference for 3D thermal reconstruction quality.

Evaluation metrics. We evaluate the quality of novel view synthesis for the RGB images using PSNR, SSIM, and LPIPS [11, 42]. PSNR quantifies pixel wise error, SSIM measures perceptual similarity between images, and LPIPS computes perceptual distance using deep features, where lower values indicate better perceptual alignment. For the TIR images, we report PSNR and SSIM,

Table 1: Quantitative comparison of TIR reconstruction performance. For space, we report PSNR, MAE_R, and MAE for each scene, while the Average columns summarize the overall performance across all scenes using all five metrics. Full per-scene results, including SSIM and LPIPS, are provided in the supplementary material. MAE_R denotes $\mathrm{MAE}_{\mathrm{ROI}}$. ↑ (↓) indicates higher (lower) is better. Best, second best, and third best results are highlighted.

Method	**Average (All Scenes)**					buildingA_spring			buildingA_winter			freezing_ice_cup			double_robot		
	PSNR↑	SSIM↑	LPIPS↓	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓
Thermo-NeRF	31.101	0.943	0.153	1.550	0.603	26.630	2.400	1.880	28.750	0.660	0.760	30.670	3.260	0.570	30.750	0.910	0.340
MFTG	29.472	0.964	0.093	1.829	0.700	25.742	1.756	2.334	26.613	0.962	1.093	27.214	0.465	0.762	29.672	2.804	0.426
MSMG	28.155	0.962	0.116	1.924	0.788	27.157	1.580	2.019	21.691	1.528	1.954	32.174	0.456	0.705	30.985	1.815	0.405
OMMG	32.197	0.970	0.084	1.148	0.490	29.130	0.921	1.325	30.969	0.545	0.618	31.033	0.286	0.432	31.878	1.441	0.309
Thermal3DGS	28.397	0.958	0.081	2.335	1.041	25.093	3.220	3.328	25.832	1.261	1.286	30.000	0.457	0.890	27.139	1.491	0.724
3DGS+IR	30.910	0.959	0.162	1.233	0.606	27.390	1.347	1.734	29.750	0.654	0.700	30.890	0.398	0.515	30.160	1.010	0.433
Ours	32.479	0.975	0.068	1.425	0.472	29.970	0.922	1.345	31.460	0.485	0.536	29.710	0.298	0.481	28.970	3.419	0.416

Method	exhibition_building			heater_water_cup			heater_water_kettle			melting_ice_cup			raspberrypi			trees		
	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓	PSNR↑	MAE_R↓	MAE↓
Thermo-NeRF	33.750	0.310	0.350	32.250	2.100	0.530	34.040	2.760	0.710	31.300	1.570	0.290	31.800	1.280	0.270	31.070	0.250	0.330
MFTG	33.063	0.380	0.408	28.780	5.468	0.478	29.679	4.076	0.655	29.679	0.223	0.325	33.713	1.907	0.230	30.560	0.246	0.293
MSMG	29.132	0.439	0.523	26.850	5.346	0.493	27.641	4.124	0.710	26.432	0.247	0.437	30.931	3.437	0.273	28.560	0.273	0.361
OMMG	34.528	0.259	0.305	31.793	4.034	0.467	35.245	2.747	0.515	27.690	0.394	0.541	37.382	0.672	0.156	32.327	0.186	0.228
Thermal3DGS	27.899	1.060	1.016	28.179	6.949	0.664	30.959	4.095	1.334	29.844	0.351	0.418	30.459	4.066	0.301	28.564	0.399	0.454
3DGS+IR	33.770	0.291	0.341	30.440	4.811	0.493	34.981	2.469	0.703	25.450	0.435	0.586	34.910	0.663	0.249	31.360	0.255	0.305
Ours	36.410	0.304	0.317	30.780	4.703	0.466	34.540	3.067	0.541	31.090	0.180	0.254	38.250	0.692	0.129	33.610	0.183	0.231

along with temperature errors to assess physical accuracy. Specifically, we compute Mean Absolute Error (MAE) [33] and region of interest MAE ($\mathrm{MAE}_{\mathrm{ROI}}$) [8] to evaluate the accuracy of temperature inversion results.

Baselines. The proposed method is compared against representative baselines [40] from thermal reconstruction and general purpose novel view synthesis, including thermal Gaussian variants (MFTG, MSMG, OMMG) [21], Thermal3DGS [3], ThermalNeRF [8], Scaffold-GS [22], and Mip-Splatting [39].

Implementation details. The proposed method is implemented in PyTorch and trained on a single NVIDIA RTX 3090 GPU. Adam is used with $\beta = (0.9, 0.999)$ for 30,000 iterations. For geometric/position parameters, the learning rate decays exponentially from 1.6×10^{-4} to 1.6×10^{-6}. For the proposed physical parameters, emissivity ϵ, temperature T, and reflectance ρ, learning rates are initialized to 0.0025 and adjusted during training. Loss weights are set to $\lambda_{\mathrm{rgb\_grad}} = 0.05$, $\lambda_{\mathrm{ir\_guided\_tv}} = 0.02$, and $\lambda_{\mathrm{ir\_physical\_prior}} = 0.01$. To ensure a stable transition from shared geometry to modality-specific appearance, we employ a two-stage optimization strategy. Following 3,000 iterations of geometric scaffolding, a Non-uniform Alternating Optimization Schedule decouples RGB and TIR opacity fields by periodically freezing specific parameter groups (covering 20% of total iterations).

4.2 Comparison Results

On the ThermoNeRF dataset, RaViGS attains superior TIR fidelity with 32.48 dB PSNR and 0.98 SSIM, markedly exceeding ThermalNeRF (Table 1). Beyond image-level metrics, our framework yields a substantially lower temperature MAE than the channel-concatenation baseline (e.g., 3DGS+IR). We attribute

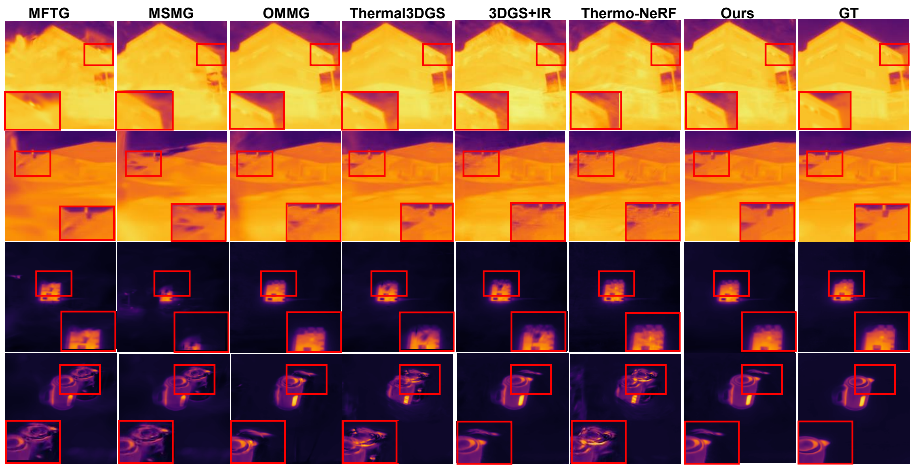

Fig. 2: Qualitative comparison of TIR novel view synthesis on the ThermoNeRF dataset. Compared to the baselines, our RaViGS successfully suppresses boundary ghosting artifacts and recovers sharper TIR structural boundaries.

this improvement to resolving the radiometric mismatch inherent, where treating TIR as an auxiliary color channel detaches observations from radiative physics and induces view-dependent temperature drift. Conversely, our Intrinsic Radiative Attribute Factorization aligns supervision with a physically interpretable decomposition of emissivity, temperature, and reflectance. Qualitative results (Fig. 2) demonstrate that RaViGS recovers sharper boundaries and smoother thermal transitions at material interfaces, validating the efficacy of the Edge-aware Smoothness Loss in regularizing cross-modal structures.

Regarding RGB rendering results (shown in Table 2), RaViGS outperforms Mip-Splatting and Scaffold-GS with an average PSNR of 22.01 dB. This performance gain validates our dual-opacity design: by decoupling modality-specific visibility within a shared geometry, high-frequency RGB textures are effectively shielded from conflicting gradients of low-frequency thermal patterns. Consequently, RaViGS successfully mitigates blur and fusion-induced artifacts, preserving intricate true-color details as visualized in Fig. 3.

4.3 Ablation Studies

To isolate the contributions of each component, ablations are conducted over *Intrinsic Radiative Attribute Factorization*, spectrally decoupled dual-opacity rendering, and the RGB-TIR joint optimization objective.

Table 3 reports TIR reconstruction under a progressive build-up from a baseline, by adding the three-parameter radiative model, the dual-opacity system, and the physical prior term. When the three-parameter model is removed and the TIR branch degenerates to direct radiance-field fitting, MAE increases, indicating that Stefan–Boltzmann–based modeling is required to constrain temper-

Table 2: Quantitative comparison of RGB reconstruction performance across different methods. ↑ higher is better, ↓ lower is better. We differentiate the best, second best and third best methods by colors.

Method	Average (All Scenes)			buildingA_spring			buildingA_winter			freezing_ice_cup			double_robot		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Thermo-NeRF	19.971	0.622	0.273	19.300	0.630	0.270	19.740	0.610	0.340	25.690	0.770	0.370	19.020	0.650	0.300
Scaffold-GS	20.077	0.625	0.340	22.141	0.789	0.239	19.462	0.637	0.364	26.685	0.853	0.370	19.489	0.690	0.325
Mip-Splatting	20.536	0.664	0.345	18.832	0.731	0.302	18.813	0.619	0.429	25.805	0.846	0.393	20.622	0.732	0.310
MFTG	20.466	0.693	0.345	21.871	0.795	0.253	19.764	0.666	0.409	26.752	0.860	0.394	21.358	0.774	0.301
MSMG	20.551	0.708	0.383	22.676	0.809	0.282	20.659	0.702	0.413	27.384	0.866	0.398	20.327	0.738	0.409
OMMG	21.622	0.706	0.358	23.015	0.818	0.251	20.713	0.700	0.395	26.753	0.864	0.404	20.371	0.738	0.349
3DGS+IR	21.878	0.710	0.341	21.940	0.802	0.250	20.360	0.682	0.396	27.280	0.865	0.382	21.700	0.784	0.320
Ours	22.011	0.713	0.337	23.510	0.816	0.246	20.940	0.696	0.383	26.430	0.860	0.391	20.950	0.767	0.330

Method	exhibition_building			heater_water_cup			heater_water_kettle			melting_ice_cup			raspberrypi			trees		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Thermo-NeRF	22.810	0.660	0.240	18.160	0.560	0.250	21.050	0.600	0.240	16.200	0.430	0.180	19.070	0.720	0.190	18.670	0.590	0.350
Scaffold-GS	20.671	0.594	0.285	16.893	0.446	0.410	22.016	0.580	0.327	14.048	0.343	0.436	21.929	0.804	0.243	17.435	0.511	0.406
Mip-Splatting	23.547	0.730	0.226	15.618	0.437	0.471	23.713	0.694	0.280	14.153	0.393	0.470	25.451	0.885	0.201	18.802	0.572	0.368
MFTG	23.138	0.727	0.238	16.870	0.475	0.417	15.066	0.710	0.325	15.066	0.436	0.444	25.609	0.890	0.202	19.162	0.602	0.463
MSMG	21.763	0.679	0.367	18.762	0.573	0.431	15.612	0.737	0.333	16.097	0.495	0.477	24.960	0.878	0.240	17.271	0.606	0.485
OMMG	23.727	0.724	0.272	18.135	0.533	0.427	24.328	0.725	0.321	15.064	0.450	0.488	25.149	0.885	0.216	18.968	0.621	0.453
3DGS+IR	24.130	0.737	0.285	18.720	0.545	0.415	24.720	0.745	0.262	15.290	0.451	0.443	25.540	0.890	0.253	19.100	0.603	0.403
Ours	24.460	0.732	0.253	17.490	0.518	0.500	24.270	0.728	0.276	16.510	0.482	0.413	25.890	0.892	0.203	19.660	0.636	0.380

Table 3: Ablation study of the proposed components for TIR and RGB reconstruction. We report the average metrics across all scenes. ↑ higher is better, ↓ lower is better. **Bold** indicates the best performance.

	Module Configuration			TIR (Average)					RGB (Average)		
	Dual-Opacity	3-Physics	Loss-Constraint	PSNR↑	SSIM↑	LPIPS↓	MAE_{ROI}(°C)↓	MAE↓	PSNR↑	SSIM↑	LPIPS↓
Baseline	×	×	×	30.910	0.959	0.162	1.233	0.606	21.878	0.710	0.341
Dual-Opacity	✓	×	×	32.068	0.969	0.075	1.194	0.517	21.358	0.693	0.359
3-Physics	×	✓	×	31.544	0.962	0.180	1.850	2.259	21.254	0.685	0.344
Loss-Constraint	×	×	✓	31.422	0.966	0.118	**1.132**	0.522	21.080	0.691	0.344
Dual-Opacity + 3-Physics	✓	✓	×	31.567	0.962	0.071	1.334	0.528	21.316	0.696	0.340
3-Physics + Loss-Constraint	×	✓	✓	31.617	0.961	0.069	1.276	0.548	21.262	0.686	0.340
Ours	✓	✓	✓	**32.479**	**0.971**	**0.068**	1.425	**0.472**	**22.011**	**0.713**	**0.337**

ature inversion. When the dual-opacity system is removed, boundary ghosting becomes visible in TIR renderings and PSNR drops by about 1.5 dB, which is consistent with the need for modality specific occlusion relationships under shared geometry. Fig. 4 visualizes the corresponding dependency trends.

Loss ablations further examine the joint objective under bi-modal optimization (Table 4). The baseline setting (w/o Any Loss) uses only standard photometric reconstruction ($\mathcal{L}_1$ and D-SSIM), yielding $\text{PSNR}_{IR} = 32.05$ dB and $\text{PSNR}_{RGB} = 21.64$ dB. Introducing the TIR physical prior loss $\mathcal{L}_{prior}$ constrains the solution space of emissivity, temperature, and reflectance, and reduces MAE_{ROI} from 1.91. Adding RGB gradient consistency $\mathcal{L}_{grad}$ and TIR guided smoothness $\mathcal{L}_{smooth}$ transfers high frequency structural cues from RGB to TIR, alleviating edge blur in thermal renderings. With all terms enabled, the reported performance reaches $\text{PSNR}_{IR} = 32.48$ dB and $\text{PSNR}_{RGB} = 22.01$ dB, with SSIM attaining the highest value under the evaluated configurations.

Table 3 and Table 4 summarize module ablations and loss ablations for RGB rendering quality. Although the primary design targets multi-modal fusion and

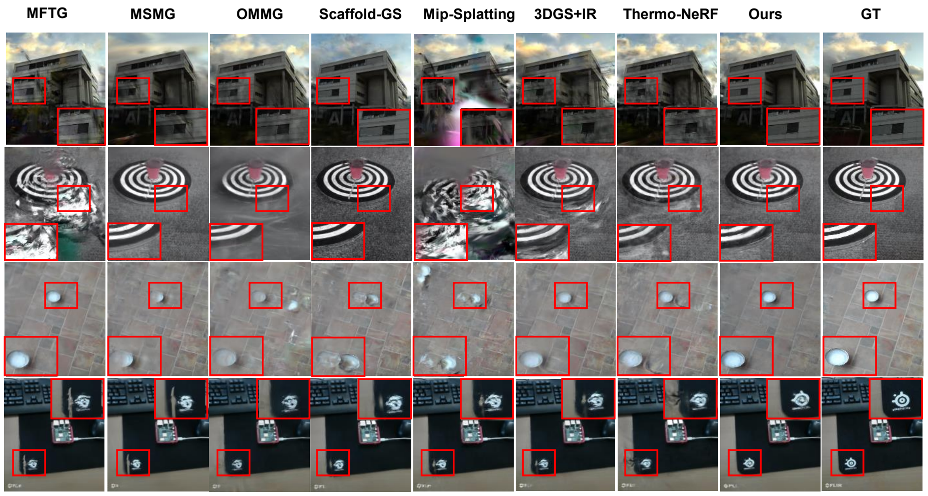

Fig. 3: Comparison of RGB novel view synthesis on the ThermoNeRF dataset. Compared to the baselines, our method successfully mitigates fusion-induced artifacts and recovers sharper high frequency true-color textures.

Table 4: Ablation study of different loss functions for TIR and RGB reconstruction. We report average metrics across all scenes. $\mathcal{L}_{grad}$, $\mathcal{L}_{gtv}$, $\mathcal{L}_{phy}$, and $\mathcal{L}_{sh}$ denote gradient loss, total variation loss, physics-based loss, and SH loss, respectively. **Bold** indicates the best performance.

	Module Configuration				TIR (Average)					RGB (Average)		
	Gradient Loss	Guided TV	Physical Prior	SH Rest L2	PSNR↑	SSIM↑	LPIPS↓	MAE_{ROI}(°C)↓	MAE↓	PSNR↑	SSIM↑	LPIPS↓
Baseline	×	×	×	×	32.050	0.967	0.162	1.912	0.497	21.641	0.704	0.341
only gradient loss	✓	×	×	×	32.402	**0.971**	0.071	1.645	0.537	21.872	0.710	0.348
only guided TV	×	✓	×	×	32.356	0.968	0.072	1.431	0.538	21.921	0.711	0.348
only physical prior	×	×	✓	×	32.456	0.969	0.069	**1.350**	0.486	22.000	0.714	0.343
only SH rest L2	×	×	×	✓	32.000	0.968	0.069	2.024	0.501	22.026	0.701	0.350
gradient loss + guided TV	✓	✓	×	×	29.661	0.969	0.069	1.413	**0.465**	21.592	0.704	0.345
gradient loss + physical prior	✓	×	✓	×	32.444	0.970	0.070	1.558	0.486	22.000	0.712	0.343
guided TV + physical prior	×	✓	✓	×	32.162	0.970	0.069	1.667	0.496	19.954	0.715	0.344
Ours	✓	✓	✓	✓	**32.479**	**0.971**	**0.068**	1.425	0.472	**22.011**	**0.713**	**0.337**

radiative modeling, the RGB branch also varies with the degree of thermal disentanglement: as thermal factors become better separated, RGB optimization is less exposed to cross modal heterogeneity. Across the full comparison set, the method retains the real time rendering characteristics of 3DGS while reporting higher reconstruction scores in both modalities.

Fig. 4 shows the qualitative ablation studies of RaViGS components. We evaluate the contributions of DO (Dual-Opacity Rendering), Physics (3-Physics Factorization), and LC (Joint Loss Constraints). Compared to the Baseline (vanilla 3DGS with channel concatenation), partial configurations exhibit silhouette ghosting (e.g., Phys+LC) or radiometrically inconsistent thermal fields (e.g., LC). Ours denotes the full RaViGS method, which successfully recovers sharp RGB details and physically consistent TIR attributes. Red insets provide magnified views for comparison against Ground Truth (GT). Despite marginal

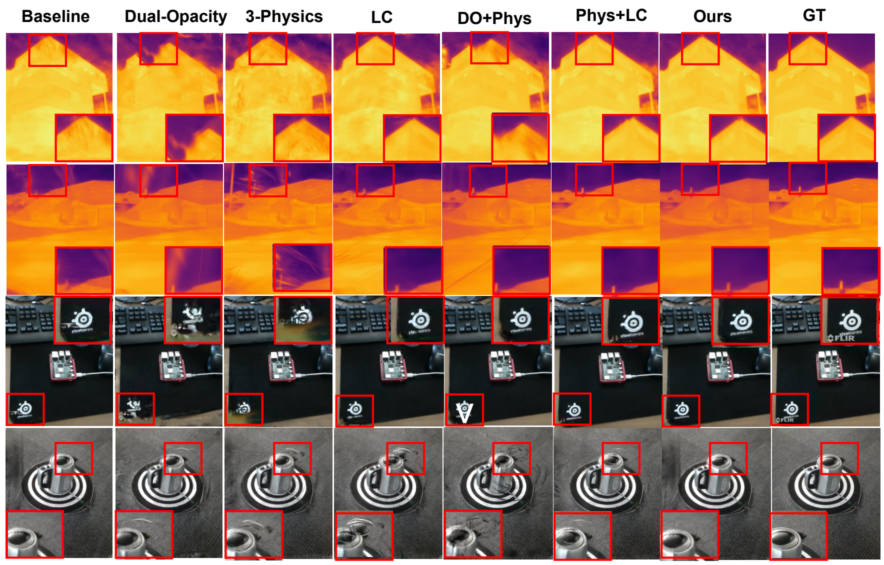

Fig. 4: Qualitative ablation results of proposed modules on joint RGB-TIR reconstruction. Compared to partial configurations, our full PhysiGS preserves sharp RGB details and physically grounded TIR consistency. Specifically, **phys+LC** (3-Physics + Loss-Constraint) induces severe silhouette ghosting. **LC** (Loss-Constraint) leads to radiometrically implausible thermal fields. **DO+Phys** (Dual-Opacity + 3-Physics) results in blurred structural boundaries.

overhead bounded by joint pruning, we observe that explicitly decoupling visibility fields effectively insulates shared geometry from thermal noise, thereby ensuring stable convergence against ambiguous radiometric factorization on specular surfaces.

Furthermore, to investigate the behavior of our shared geometry and co-support pruning strategy under extreme cross-spectral conditions, we evaluate regions where RGB and TIR exhibit fundamentally different visibility patterns, such as glass. While one might assume co-support pruning would mistakenly eliminate such regions, practical transparent surfaces usually retain weak but non-zero RGB cues, such as structural edges, specular highlights, or reflections. With our low pruning threshold ($\tau_\alpha = 0.005$), the primitives representing these details successfully survive the pruning process. As shown in Fig. 5, in glass regions, the optimized α^{rgb} remains appropriately low but non-zero, while α^{ir} converges to a clearly higher value, ensuring the relevant primitives are retained. This indicates that the co-support criterion does not destroy the geometric representation of such objects; instead, the dual-opacity mechanism effectively captures the divergent visibility on the shared scaffold. Nevertheless, we acknowledge that fully textureless transparent surfaces entirely devoid of any RGB evidence represent an extreme edge case that remains a limitation of the current shared-geometry paradigm.

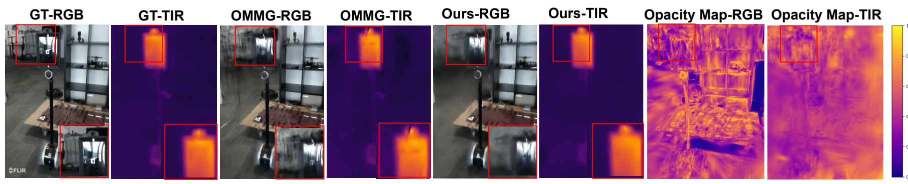

Fig. 5: Glass cases under divergent cross-spectral visibility. Practical transparent surfaces exhibit weak RGB cues (e.g., edges or reflections) that allow relevant primitives to survive co-support pruning. Our dual-opacity mechanism successfully models a low α^{rgb} and a high α^{ir} simultaneously on the shared geometric scaffold.

5 Conclusion

We present RaViGS, a physically grounded approach to joint RGB-TIR reconstruction that addresses the fundamental incompatibility between reflected and emitted radiance within a unified 3D representation. By factorizing thermal response into emissivity, temperature, and reflectance, we enable physically interpretable thermal field estimation. The dual-opacity design resolves cross-spectral visibility disparities while maintaining geometric coherence, allowing high-frequency RGB structure and low-frequency thermal attributes to complement rather than interfere. Our framework currently assumes opaque surfaces and static scenes; extending to translucent materials and dynamic thermal evolution remains future work. Nevertheless, RaViGS establishes a foundation for physically principled neural radiance fields in multi-spectral imaging, with potential applications in night-time navigation, industrial inspection, and thermographic analysis.

Acknowledgements

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No.42501545) and the Young Scientists Fund of the National Natural Science Foundation of Zhejiang Province (Grant No.QN26D010023). It was supported in part by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No.62506103) and the Fundamental Research Funds for the Provincial Universities of Zhejiang (Grant No. GK259909299001-026).

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 5855–5864 (2021)

2. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5470–5479 (2022)
3. Chen, Q., Shu, S., Bai, X.: Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis. In: European Conference on Computer Vision. pp. 253–269. Springer (2024)
4. Cheng, K., Long, X., Yang, K., Yao, Y., Yin, W., Ma, Y., Wang, W., Chen, X.: Gaussianpro: 3d gaussian splatting with progressive propagation. In: Forty-first International Conference on Machine Learning (2024)
5. Conte, P., Girelli, V.A., Mandanici, E.: Structure from motion for aerial thermal imagery at city scale: Pre-processing, camera calibration, accuracy assessment. ISPRS journal of photogrammetry and remote sensing **146**, 320–333 (2018)
6. Fox, M., Goodhew, S., De Wilde, P.: Building defect detection: External versus internal thermography. Building and Environment **105**, 317–331 (2016)
7. Gade, R., Moeslund, T.B.: Thermal cameras and applications: a survey. Machine vision and applications **25**(1), 245–262 (2014)
8. Hassan, M., Forest, F., Fink, O., Mielle, M.: Thermonerf: Joint rgb and thermal novel view synthesis for building facades using multimodal neural radiance fields. arXiv preprint arXiv:2403.12154 (2024)
9. Hassan, M., Forest, F., Fink, O., Mielle, M.: Thermonerf: A multimodal neural radiance field for joint rgb-thermal novel view synthesis of building facades. Advanced Engineering Informatics **65**, 103345 (2025)
10. He, L., Li, L., Sun, W., Han, Z., Liu, Y., Zheng, S., Wang, J., Li, K.: Neural radiance field in autonomous driving: A survey. arXiv preprint arXiv:2404.13816 (2024)
11. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
12. Huang, B., Yu, Z., Chen, A., Geiger, A., Gao, S.: 2d gaussian splatting for geometrically accurate radiance fields. In: ACM SIGGRAPH 2024 conference papers. pp. 1–11 (2024)
13. Huang, X., Zhang, Q., Feng, Y., Li, H., Wang, X., Wang, Q.: Hdr-nerf: High dynamic range neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18398–18408 (2022)
14. Hwang, S., Park, J., Kim, N., Choi, Y., So Kweon, I.: Multispectral pedestrian detection: Benchmark dataset and baseline. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1037–1045 (2015)
15. Jin, H., Liu, I., Xu, P., Zhang, X., Han, S., Bi, S., Zhou, X., Xu, Z., Su, H.: Tensoir: Tensorial inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 165–174 (2023)
16. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., et al.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
17. Liang, Z., Zhang, Q., Feng, Y., Shan, Y., Jia, K.: Gs-ir: 3d gaussian splatting for inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21644–21653 (2024)
18. Lin, D., Jarzabek-Rychard, M., Tong, X., Maas, H.G.: Fusion of thermal imagery with point clouds for building façade thermal attribute mapping. ISPRS journal of photogrammetry and remote sensing **151**, 162–175 (2019)
19. Lin, Y.Y., Pan, X.Y., Fridovich-Keil, S., Wetzstein, G.: Thermalnerf: Thermal radiance fields. In: 2024 IEEE International Conference on Computational Photography (ICCP). pp. 1–12. IEEE (2024)

20. Liu, D., Wang, Z., Chen, P.: Dsem-nerf: Multimodal feature fusion and global–local attention for enhanced 3d scene reconstruction. Information Fusion **115**, 102752 (2025)
21. Lu, R., Chen, H., Zhu, Z., Qin, Y., Lu, M., Zhang, L., Yan, C., Xue, A.: Thermalgaussian: Thermal 3d gaussian splatting. arXiv preprint arXiv:2409.07200 (2024)
22. Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., Dai, B.: Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20654–20664 (2024)
23. Maset, E., Fusiello, A., Crosilla, F., Toldo, R., Zorzetto, D., et al.: Photogrammetric 3d building reconstruction from thermal images. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **4**, 25–32 (2017)
24. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
25. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM transactions on graphics (TOG) **41**(4), 1–15 (2022)
26. Planinsic, G.: Infrared thermal imaging: Fundamentals, research and applications. European Journal of Physics **32**(5), 1431 (2011)
27. Poggi, M., Ramirez, P.Z., Tosi, F., Salti, S., Mattoccia, S., Di Stefano, L.: Cross-spectral neural radiance fields. In: 2022 International Conference on 3D Vision (3DV). pp. 606–616. IEEE (2022)
28. Qin, M., Li, W., Zhou, J., Wang, H., Pfister, H.: Langsplat: 3d language gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20051–20060 (2024)
29. Thirgood, C., Mendez, O., Ling, E., Storey, J., Hadfield, S.: Hypergs: Hyperspectral 3d gaussian splatting. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5970–5979 (2025)
30. Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P.: Ref-nerf: Structured view-dependent appearance for neural radiance fields. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(11), 9426–9437 (2024)
31. Vogiatzis, G., Hernández, C.: Video-based, real-time multi-view stereo. Image and vision computing **29**(7), 434–441 (2011)
32. Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M.: 'structure-from-motion'photogrammetry: A low-cost, effective tool for geoscience applications. Geomorphology **179**, 300–314 (2012)
33. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. Climate research **30**(1), 79–82 (2005)
34. Xie, T., Zong, Z., Qiu, Y., Li, X., Feng, Y., Yang, Y., Jiang, C.: Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4389–4398 (2024)
35. Yan, Y., Lin, H., Zhou, C., Wang, W., Sun, H., Zhan, K., Lang, X., Zhou, X., Peng, S.: Street gaussians for modeling dynamic urban scenes.(2023). arXiv preprint arXiv:2401.01339 (2023)
36. Yan, Z., Low, W.F., Chen, Y., Lee, G.H.: Multi-scale 3d gaussian splatting for anti-aliased rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20923–20931 (2024)

37. Yang, K., Liu, Y., Cui, Z., Liu, Y., Zhang, M., Yan, S., Wang, Q.: Ntr-gaussian: Nighttime dynamic thermal reconstruction with 4d gaussian splatting based on thermodynamics. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 691–700 (2025)
38. Ye, T., Wu, Q., Deng, J., Liu, G., Liu, L., Xia, S., Pang, L., Yu, W., Pei, L.: Thermal-nerf: Neural radiance fields from an infrared camera. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 1046–1053. IEEE (2024)
39. Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-splatting: Alias-free 3d gaussian splatting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19447–19456 (2024)
40. Zhang, H., Yuan, D., Shu, X., Li, Z., Liu, Q., Chang, X., He, Z., Shi, G.: A comprehensive review of rgbt tracking. IEEE Transactions on Instrumentation and Measurement **73**, 1–23 (2024)
41. Zhang, J., Zhan, F., Xu, M., Lu, S., Xing, E.: Fregs: 3d gaussian splatting with progressive frequency regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21424–21433 (2024)
42. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
43. Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Transactions on Graphics (ToG) **40**(6), 1–18 (2021)
44. Zhou, S., Chang, H., Jiang, S., Fan, Z., Zhu, Z., Xu, D., Chari, P., You, S., Wang, Z., Kadambi, A.: Feature 3dgs: Supercharging 3d gaussian splatting to enable distilled feature fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21676–21685 (2024)
45. Zhou, X., Lin, Z., Shan, X., Wang, Y., Sun, D., Yang, M.H.: Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 21634–21643 (2024)
46. Zhu, H., Sun, Y., Liu, C., Xia, L., Luo, J., Qiao, N., Nevatia, R., Kuo, C.H.: Multimodal neural radiance field (2023)