

PhysIR-Splat: Physically Consistent Thermal Infrared Radiative Transfer in 3D Gaussian Splatting

Jingyuan Gao¹, Yumeng Hu¹, Fei Gao¹, Mingjin Zhang^{1*}
¹Xidian University

Abstract

Thermal infrared (TIR) 3D reconstruction provides geometry that is intrinsically coupled to the temperature field, even in low-light, nighttime, and smoke-obscured environments. TIR imaging measures self-emitted thermal radiation driven by object temperature and is largely independent of external illumination; therefore, simply carrying over visible-spectrum assumptions to TIR-based 3D reconstruction and novel view synthesis (NVS) often results in floating artifacts and blurred edges. In addition, radiometric inconsistency and low contrast in TIR weaken structure-from-motion (SfM) initialization, which in turn hinders subsequent 3D Gaussian Splatting (3DGS) optimization. We present **PhysIR-Splat**, a 3DGS framework that follows infrared radiative transfer: we explicitly model temperature, emissivity, and environmental irradiance on Gaussian primitives and, during rendering, jointly account for thermal emission, the reflected component, and atmospheric transmittance to produce physically consistent thermal synthesis. We also introduce **VGGT-IR**, a Transformer-based feed-forward initializer that takes TIR input with optional RGB and directly regresses camera poses and initial geometry, providing a modality-aligned and stable starting point for PhysIR-Splat. Extensive experiments demonstrate that our method significantly surpasses existing approaches in thermal reconstruction quality and cross-view consistency, effectively suppressing floating artifacts and enhancing boundary sharpness. The source code will be publicly available at <https://github.com/JingyuanGao0919/physir-splat>.

1. Introduction

Thermal imaging senses object temperature by measuring self-emitted radiation without contact. For multi-view *thermal infrared* (TIR) sequences, 3D thermal reconstruction aims to recover surface radiance and temperature distributions, with broad applications in healthcare, emerging med-

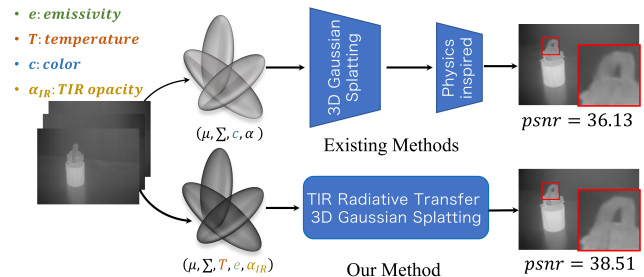


Figure 1. Prior work treats 3DGS color as emitted radiance and learns a physics-inspired module (e.g., ATF/TCM [7]) to correct the renderings, whereas we assign physical quantities to 3DGS primitives and embed Thermal-IR radiative transfer directly into the renderer.

ical digital-twin and metaverse systems, industrial inspection, agriculture, and building diagnostics [8, 11, 15, 27, 56]. Conventional approaches typically (i) hand-craft scene geometry and material parameters and synthesize TIR images using the *radiative transfer equation* (RTE) [3, 29, 37, 50]; or (ii) first reconstruct 3D shape from multi-view RGB and then texture-map the thermal images onto the recovered geometry [28, 40, 55]. The former struggles to obtain material and environmental descriptions faithful to reality, whereas the latter faces challenging cross-spectral registration and tends to produce streaking/ghosting (“floaters”) and edge tearing.

Recent *3D Gaussian Splatting* (3DGS) [25] has advanced *novel view synthesis* (NVS) and 3D reconstruction, and has been quickly adapted to thermal imaging. VetaGS [35] employs a view-dependent deformation field driven by camera pose and viewing direction to capture thermal variation; ThermalGaussian [30] adopts a multi-modal Gaussian strategy; Thermal3D-GS [7] models atmospheric transfer and heat conduction to improve 3DGS rendering for TIR, boosting novel view synthesis quality. However, many existing methods retrofit visible-light assumptions and directly treat the 3DGS color as thermal radiance, as illustrated in Fig. 1. Meanwhile, the low texture and radiometric inconsistency of TIR imagery render keypoint detection

*Corresponding author. Email: mjinzhang@xidian.edu.cn

and matching unreliable, so *Structure-from-Motion* (SfM)-based [41] pose estimation often degrades, which in turn weakens 3DGS initialization and convergence.

We propose **PhysIR-Splat**, a physically grounded 3D Gaussian framework for thermal reconstruction. For each Gaussian, we explicitly model temperature, emissivity, and environmental irradiance. We align to the camera passband using a passband Planck integral lookup table (LUT), and represent global environmental irradiance with low-order spherical harmonics. Along each ray, we use emission-absorption volume rendering with front-to-back accumulation: each Gaussian contributes emitted radiance weighted by opacity derived from optical thickness. A small monotonic response module then fits the sensor’s radiance-to-gray mapping nonlinearity, unifying the imaging chain of *self-emission + environmental reflection* $\rightarrow$ *atmospheric transmittance* $\rightarrow$ *radiometric response*. This improves temperature fidelity and physical consistency under novel views. In parallel, we design **VGGT-IR**, which accepts TIR with optional RGB. We extract modality-specific features using DINOv3[43], fuse them additively, and use alternating attention to aggregate cross-view information. A single forward pass directly regresses camera parameters and initial geometry. By bypassing feature-based SfM pipelines that rely on keypoint detection and matching, VGGT-IR effectively mitigates the degradation caused by low texture and “ghosting” in TIR.

VGGT-IR provides robust poses and geometry, while PhysIR-Splat delivers physically consistent radiative modeling and high-fidelity rendering. The two components jointly address the dual bottlenecks of physical realism in rendering and robustness in initialization. In summary, the main contributions are as follows:

- We introduce PhysIR-Splat, a physics-driven 3D Gaussian framework that explicitly models temperature, emissivity, and environmental irradiance per Gaussian, enabling physically consistent, end-to-end learnable, and efficient thermal rendering.
- We present VGGT-IR, a Transformer-based feed-forward initializer that unifies TIR and optional RGB inputs to directly regress camera parameters and scene geometry, substantially improving initialization in low-texture and ghosting scenarios and integrating seamlessly with the 3DGS pipeline.
- We conduct comprehensive evaluations on public datasets, showing superior camera pose estimation and novel view rendering quality compared to existing methods, and validating the effectiveness and robustness of each component via ablations.

2. Related Work

Conventional 3D Thermal Reconstruction Conventional pipelines often follow a “geometry first, thermal mapping

later” paradigm: geometry is obtained from active depth (LiDAR, depth cameras, or their fusion), and TIR is then projected onto surfaces [4, 5, 26, 42, 45, 57, 58]. This route is highly sensitive to cross-sensor calibration and temporal synchronization, and further constrained by working range and reflectance. To reduce dependence on specialized hardware, photogrammetric methods perform RGB-based SfM/MVS and then register to TIR [10, 16, 20, 32, 45], but still require dense, high-quality RGB/TIR pairs. Other works map thermal images onto textureless polyhedral models [19, 24] using navigation data and extrinsics, but remain prone to insufficient matches and pose errors.

Methods based on *Neural Radiance Fields* (NeRF) [33] offer recent advances: ThermoNeRF [18] jointly renders RGB and TIR, using RGB to guide density while suppressing cross-modal leakage; Thermal-NeRF [52] explicitly models thermal values with structured thermal constraints; Overall, while these methods have advanced both geometry and thermal appearance, they still rely on precise cross-modal registration or densely paired data. Their cascaded, non-end-to-end pipelines are prone to error accumulation, and their physical consistency and robustness remain limited in low-texture scenes, under calibration bias, or under complex radiative transfer conditions.

3DGS-based Thermal Reconstruction *3D Gaussian Splatting* (3DGS) [25], with efficient differentiable rendering, has been rapidly extended to TIR. Veta-GS [35] employs view-cone masks and a view-dependent deformation field driven by pose and viewing direction, together with thermal features and frequency/edge consistency losses. ThermalGaussian [30] builds a multimodal 3DGS with multimodal initialization, IR-specific Gaussians, and cross-modal regularization to jointly reconstruct thermal and RGB while balancing quality, speed, and memory. Thermal3D-GS [7] designs physics-informed modules—incorporating priors such as atmospheric transmittance and heat conduction—together with a temperature-consistency loss to optimize 3DGS rendering, thereby improving novel view synthesis quality. Nevertheless, most 3DGS thermal extensions retrofit visible-light assumptions, lacking accurate modeling of view-dependent thermal radiation and atmospheric transmittance, which limits agreement with the true thermal field. Moreover, *non-uniform correction* (NUC) in thermal imaging hampers feature extraction and matching for SfM, making robust pose estimation challenging in 3DGS-style pipelines.

3. Method

3.1. Preliminaries

Infrared Image Decomposition. TeS [14] decomposes thermal signals into temperature T and emissivity e . TeX [2] further estimates T, e and thermal texture X

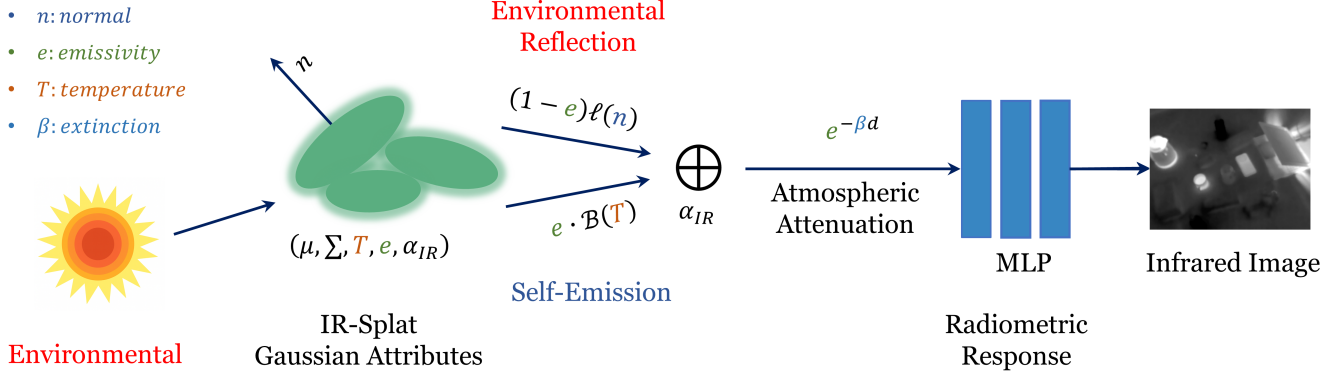


Figure 2. **PhysIR-Splat rendering pipeline.** Each Gaussian primitive carries physical attributes. The TIR imaging chain follows *self-emission + environmental reflection* $\rightarrow$ *atmospheric transmittance* $\rightarrow$ *radiometric response*.

from multi-band IR. As most datasets provide passband-integrated images and, under a *quasi-constant emissivity* assumption (for common materials, e_λ is nearly constant within the working band), PID [31] proposes the *TeV decomposition*: for pixel α ,

$$S_\alpha = e_\alpha \int_{\lambda_{\min}}^{\lambda_{\max}} B_\lambda(T_\alpha) d\lambda + (1 - e_\alpha) \int_{\lambda_{\min}}^{\lambda_{\max}} V_\lambda^\alpha \hat{S}^\lambda d\lambda, \quad (1)$$

where the first term is selfemission and the second term is reflected environmental radiance; S_α denotes the passband-integrated radiance at pixel α . The Planck spectral radiance is

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \cdot \frac{1}{\exp(\frac{hc}{\lambda kT}) - 1}, \quad (2)$$

with h (Planck), k (Boltzmann), c (light speed), and temperature T .

VGGT. VGGT [46] is a large Transformer that, in a single forward pass, regresses key 3D quantities from as few as one view to over a hundred, dispensing with traditional keypoint-based SfM pipelines. For images $\{I_i\}_{i=1}^N$,

$$f(\{I_i\}_{i=1}^N) = \{(g_i, D_i, P_i, T_i)\}_{i=1}^N, \quad (3)$$

where g_i are camera parameters (frame 1 is the world reference), D_i depth, P_i dense 3D points, and T_i any-point tracking features.

3.2. Thermal Gaussian: PhysIR-Splat

Building on the *TeV* decomposition [31], we embed the infrared image-formation chain—self-emission and environmental reflection, atmospheric transmittance, and radiometric response—into 3DGS, as illustrated in Fig. 2. We assume a clear propagation path, low aerosols, negligible LWIR/FIR scattering, and mainly diffuse targets [22]. We model self-emission with passband-integrated Planck radiance, approximate environmental irradiance by convolving

a Lambertian cosine kernel with low-order *spherical harmonics* (SH) [38], and convert the result to exitant radiance, yielding physically consistent and computationally efficient infrared shading. Geometry is shared with the RGB branch: each Gaussian’s (μ_i, Σ_i) and image-plane covariance $\Sigma_{i,\text{img}} = J_\pi \Sigma_i J_\pi^\top$ are unchanged; only radiometry and opacity are TIR-specific.

Self-emission. Self-emission is modeled as the band-integrated Planck radiance [31]:

$$\mathcal{B}(T) = \int_{\lambda_{\min}}^{\lambda_{\max}} B_\lambda(T) d\lambda, \quad (4)$$

where the integral is taken over the TIR camera’s passband. We precompute a LUT of $\mathcal{B}(T)$ and its derivatives to ensure end-to-end differentiability and numerical stability.

Environmental Reflection. We adopt *spectral-angular separability* [38]. Let the passband environment radiance be $\bar{L}_{\text{env}}(\omega) = \int_{\lambda_{\min}}^{\lambda_{\max}} V_\lambda^\alpha \hat{S}^\lambda(\omega) d\lambda \approx \sum_{l,m} c_{lm} Y_{lm}(\omega)$. Convolution with the cosine kernel with SH gives

$$E_{\text{irr}}(n) \approx \sum_{l \leq 2, m} k_l c_{lm} Y_{lm}(n), \quad k_0 = \pi, \quad k_1 = \frac{2\pi}{3}, \quad k_2 = \frac{\pi}{4}. \quad (5)$$

Define reduced irradiance $\mathcal{I}(n) = E_{\text{irr}}(n)/\pi$:

$$\mathcal{I}(n) \approx \sum_{l \leq 2, m} \bar{k}_l c_{lm} Y_{lm}(n), \quad \bar{k}_0 = 1, \quad \bar{k}_1 = \frac{2}{3}, \quad \bar{k}_2 = \frac{1}{4}. \quad (6)$$

Per-primitive Radiance. We parameterize each primitive with an in-band emissivity $e \in [0, 1]$ and set its reflectivity to $r = 1 - e$, so that the pixel-level radiance is

$$S \approx e \mathcal{B}(T) + (1 - e) \mathcal{I}(n), \quad (7)$$

As 3DGS Gaussians tend to be surfel-like, we define per-primitive radiance:

$$L_i = e_i \mathcal{B}(T_i) + (1 - e_i) \mathcal{I}(n_i), \quad (8)$$

with $T_i = \text{softplus}(\cdot) + T_{\min}$ and $e_i = \sigma(\cdot)$. The normal n_i is initialized via Normal-GS [49] (smallest-eigenvalue eigenvector of Σ_i), then optimized under shading supervision.

TIR-specific Opacity. We introduce a dedicated TIR opacity, independent of RGB. Let the unit-peak density be

$$\rho_i(\mathbf{x}) = \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^\top \Sigma_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right]. \quad (9)$$

Its line integral along $\mathbf{x}(t) = \mathbf{o} + t\mathbf{d}$ yields $\tau_i(u) = s_i w_i(u)$, where $A_i = \mathbf{d}^\top \Sigma_i^{-1} \mathbf{d}$ and $\sqrt{2\pi/A_i}$ is the equivalent thickness. The footprint $w_i(u) = \exp[-\frac{1}{2}\Delta u^\top \Sigma_{i,\text{img}}^{-1} \Delta u]$, with $\Delta u = u - \pi(\boldsymbol{\mu}_i)$, encodes ray-primitive overlap. Following geometric normalization [47], we set a scene-constant equivalent extinction. Given target center opacity $\alpha_\star \in (0, 1)$, compute $\langle \sqrt{2\pi/A_i} \rangle_i$ over current rays and define

$$\bar{\kappa} = \frac{-\ln(1-\alpha_\star)}{\langle \sqrt{2\pi/A_i} \rangle_i}, \quad s_i = \bar{\kappa} \sqrt{\frac{2\pi}{A_i}}. \quad (10)$$

Then

$$\alpha_i^{\text{IR}}(u) = 1 - \exp[-s_i w_i(u)], \quad (11)$$

used only for TIR; geometry and RGB colors are shared, while TIR uses (T_i, e_i) and α_i^{IR} .

Atmospheric Transmittance and Rendering. For a homogeneous atmosphere, transmittance is $T_{\text{atm}}(d_i) = \exp(-\beta d_i)$ with learnable extinction coefficient β and primitive-camera distance d_i . The passband radiance at pixel u is

$$\begin{aligned} \tilde{S}(u) &= \sum_{i=1}^K T_{i-1}^{\text{acc}} \alpha_i^{\text{IR}}(u) T_{\text{atm}}(d_i) L_i, \\ T_{i-1}^{\text{acc}} &= \prod_{j<i} (1 - \alpha_j^{\text{IR}}(u)), \end{aligned} \quad (12)$$

which cleanly separates local absorption, path transmittance, and emission.

Radiometric Response. Full radiative transfer includes weak near-ground path emission L_{path} , which we absorb on the response side via a global bias b and a monotonic approximator ψ_θ (three-layer narrow MLP with Softplus; weights constrained via $W_\ell = \text{softplus}(\tilde{W}_\ell)$):

$$\hat{Y}(u) = \psi_\theta(g_s \tilde{S}(u) + b), \quad (13)$$

ensuring a strictly monotonic radiance-to-gray mapping.

3.3. Thermal 3D Reconstruction: VGGT-IR

VGGT-IR targets SfM degradation on TIR sequences (Fig. 3). The input is N IR views with optional RGB from the same viewpoints. The model is IR-centric with early feature-level fusion when RGB is present:

$$f_{\text{VGGT-IR}}(\{I_i^{\text{IR}}\}_{i=1}^N, [\{I_i^{\text{RGB}}\}_{i=1}^N]) = \{(g_i, D_i, P_i)\}_{i=1}^N. \quad (14)$$

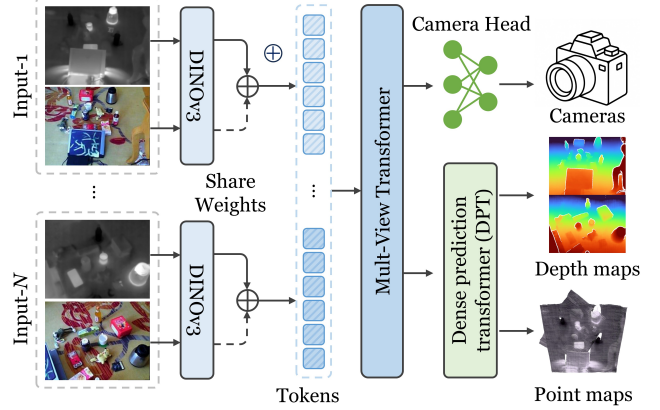


Figure 3. **Overview of VGGT-IR.** Multi-view TIR (with optional RGB) is encoded by a multimodal front-end and an alternating-attention backbone to directly regress camera poses, depth, and point maps.

Cameras $g_i = [q_i, t_i, f_i]$ use a unit-norm quaternion $q_i \in \mathbb{R}^4$, translation $t_i \in \mathbb{R}^3$, and focal $f_i = (f_{x,i}, f_{y,i})$ with the principal point fixed at the image center. For pixels

$$\mathcal{P}(I_i) = \{1, \dots, H\} \times \{1, \dots, W\}, \quad (15)$$

$D_i : \mathcal{P}(I_i) \rightarrow \mathbb{R}_+$ gives depth and $P_i : \mathcal{P}(I_i) \rightarrow \mathbb{R}^3$ maps to world points.

We map IR (with optional RGB) into a shared latent space via separate distilled DINOv3 ViT-L/16 encoders [43]. We adapt our backbone by taking the average of the three channels in the weights of the patch embed module and using it as the weight for the IR channel. We extract the patch embeddings after the final layer normalization and denote them, for view i , by

$$F_{\text{IR}} \in \mathbb{R}^{1024 \times \frac{H}{16} \times \frac{W}{16}}, \quad [F_{\text{RGB}} \in \mathbb{R}^{1024 \times \frac{H}{16} \times \frac{W}{16}}], \quad (16)$$

apply layer normalization (LN) separately to the features from the two modalities and sum them (falling back to IR only when RGB is unavailable). A second LN yields the per-view encoding E_i , which is flattened into tokens, denoted $F_E^{(i)} \in \mathbb{R}^{1024 \times \frac{HW}{256}}$. A constant reference embedding marks frame 1. Multi-view interaction uses a 12-layer alternating-attention Transformer [46] (12 heads, width 1024, MLP ratio 4).

Image tokens ($F_E^{(i)}$) are passed through a *Dense Prediction Transformer* (DPT)[39] block to produce a dense feature map. Two lightweight heads then yield the camera and dense predictions: (i) a stack of four attention layers followed by a linear regressor predicts $\mathbf{g}_i = [\mathbf{q}_i, \mathbf{t}_i, f_i]$ (with $\mathbf{q}_i$ normalized to unit norm); (ii) a 3×3 convolutional head outputs the depth map D_i and the point map P_i .

We then run a lightweight bundle adjustment (BA) [17] minimizing reprojection error to refine $\{g_i\}$ and a well-

observed subset $\mathcal{X}_{\text{sub}}$:

$$\{\hat{g}_i\}, \hat{\mathcal{X}}_{\text{sub}} = \arg \min_{\{g_i\}, \mathcal{X}_{\text{sub}}} \sum_i \sum_{\mathbf{y} \in \Omega_i} \rho(\|\pi(g_i, \mathbf{X}_{\mathbf{y}}) - \mathbf{y}\|_2^2), \quad (17)$$

with projection $\pi(\cdot)$, robust penalty $\rho(\cdot)$, and valid-pixel set Ω_i . The resulting $\{\hat{g}_i\}$ and $\hat{\mathcal{X}}_{\text{sub}}$ initialize the 3DGS stage.

3.4. Training

Thermal Loss. The total loss is

$$\mathcal{L}_{\text{TIR}} = (1 - \lambda - \lambda_{\text{smooth}} - \lambda_{\text{emis}}) \mathcal{L}_1 + \lambda \mathcal{L}_{\text{SSIM}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}} + \lambda_{\text{emis}} \mathcal{L}_{\text{emis}}. \quad (18)$$

Smoothness is enforced as a 4-neighborhood total-variation penalty on the post-response image Y :

$$\mathcal{L}_{\text{smooth}} = \frac{1}{4M} \sum_{(i,j) \in \Omega} (|Y_{i\pm 1,j} - Y_{i,j}| + |Y_{i,j\pm 1} - Y_{i,j}|), \quad (19)$$

with pixel set Ω and $M = |\Omega|$. The emissivity-sparse-reflection prior is

$$\mathcal{L}_{\text{emis}} = \frac{1}{N} \sum_{k=1}^N \phi([r_k - r_0]_+), \quad r_k = 1 - e_k, \quad (20)$$

where $e_k \in (0, 1)$ is the k -th primitive emissivity, $r_0 > 0$ a small tolerance, N the number of primitives, $[x]_+ = \max(0, x)$, and ϕ the Huber penalty [21].

Training VGGT-IR. We train VGGT-IR end-to-end with the VGGT loss design. To handle *optional RGB*, we add RGB with probability 0.2 during training (otherwise IR only), keeping the model IR-centric yet able to exploit RGB. Paired IR-visible samples are built by feeding RGB images from BlendedMVS[51], ScanNet++ v2[53], and Mapillary Planet-Scale Depth[1] into a PID-based [31] diffusion generator to synthesize pixel-aligned IR while preserving geometry and parallax. We also include real multi-modal data from the Multi-Spectral Dataset [9], which provides synchronized RGB, LWIR, depth, and pose annotations, using the depth and poses to construct point maps.

4. EXPERIMENT

4.1. Infrared 3D Reconstruction: VGGT-IR

We systematically evaluate the 3D reconstruction performance of VGGT-IR, VGGT [46], and COLMAP [41] on the LWIR sequences of the Multi-Spectral Dataset [9]. The dataset is captured with a synchronized RGB camera, an LWIR camera, and a Kinect v2, and provides ground-truth camera trajectories under diverse indoor/outdoor illumination conditions. All methods run with unified hyperparameters on a single NVIDIA H20 (96 GB): we fix the

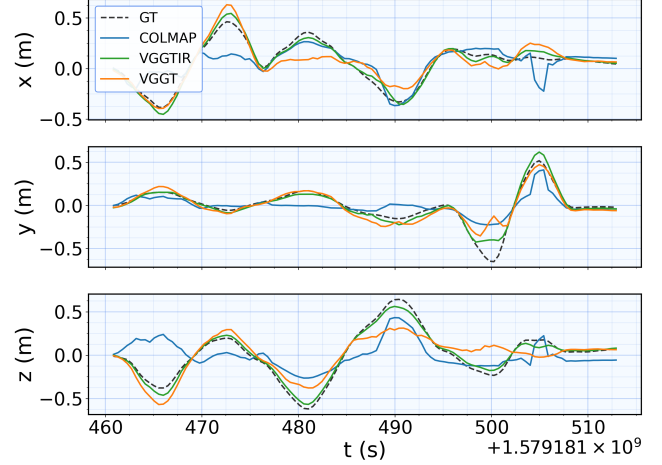


Figure 4. Camera pose estimates of COLMAP, VGGT, and VGGT-IR compared with the ground-truth trajectory on the desk1-xyz-dim2 dataset. The horizontal axis denotes time t , and the vertical axis shows the x , y , and z coordinate components.

Table 1. Root-mean-square translation error (RMSE) of Absolute Pose Error (APE $\downarrow$) on the Multi-Spectral dataset [9].

Sequence	COLMAP	VGGT	VGGT-IR
desk1-xyz-dim2	0.3107	0.2133	0.0823
desk1-halfsphere	0.1278	0.0922	0.0807
desk2-halfsphere-ic	0.0643	0.0472	0.0176

RANSAC inlier threshold, perform full bundle adjustment (BA) for each sequence, and then align the estimated poses to the ground truth using the Umeyama similarity transform (global rotation $\mathbf{R}$, translation $\mathbf{t}$, scale s). The metric is the aligned Absolute Pose Error (APE)-Trans (m), computed from aligned camera centers.

As shown in Fig. 4, on *desk1-xyz-dim2*, COLMAP drifts in weak-texture, low-contrast TIR images; VGGT alleviates this slightly, but long sequences still accumulate errors. VGGT-IR stays closer to the ground truth along the whole trajectory, with more stable poses. The quantitative results in Tab. 1 also show that VGGT-IR achieves lower APE-Trans (m) on all three sequences. In summary, VGGT-IR is more adaptive to low texture, low contrast, and NUC residuals, producing more robust poses and reconstructions.

4.2. Thermal Gaussian: PhysIR-Splat

Setup. Our PhysIR-Splat is implemented in PyTorch and trained on a single RTX 4090 (24 GB). We use Adam with cosine annealing, decaying the learning rate to 1.6×10^{-6} ; each scene is trained for 30k iterations; every 8th image is used for testing. The metrics are PSNR ($\uparrow$), SSIM [48] ($\uparrow$), and LPIPS [54] ($\downarrow$). The IR channel’s center opacity target is set to $\alpha_* = 0.6$ (indoor) and $\alpha_* = 0.4$ (outdoor). We set $\lambda = 0.2$, $\lambda_{\text{smooth}} = 0.05$, and $\lambda_{\text{emis}} = 0.05$. Most existing

Table 2. Quantitative results with state-of-the-art IR-only methods on the TI-NSD datasets [7].

Method	Indoor			Outdoor			UAV			Average		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
InstantNGP [34]	27.46	0.918	0.289	26.45	0.839	0.292	20.82	0.680	0.387	24.91	0.812	0.323
3DGS [25]	32.98	0.953	0.259	28.89	0.904	0.227	34.51	0.953	0.119	32.01	0.936	0.206
Thermal3D-GS [7]	36.01	0.962	0.252	32.60	0.942	0.187	36.74	0.962	0.112	35.04	0.955	0.187
Veta-GS [35]	37.07	0.963	0.240	33.78	0.947	0.168	37.05	0.963	0.100	35.97	0.958	0.169
Ours	<u>37.83</u>	<u>0.964</u>	<u>0.236</u>	<u>34.25</u>	<u>0.950</u>	<u>0.138</u>	<u>37.46</u>	<u>0.967</u>	<u>0.098</u>	<u>36.47</u>	<u>0.960</u>	<u>0.158</u>
Ours(VGGT-IR)	38.41	0.967	0.228	35.57	0.952	0.126	38.12	0.969	0.096	37.33	0.962	0.150

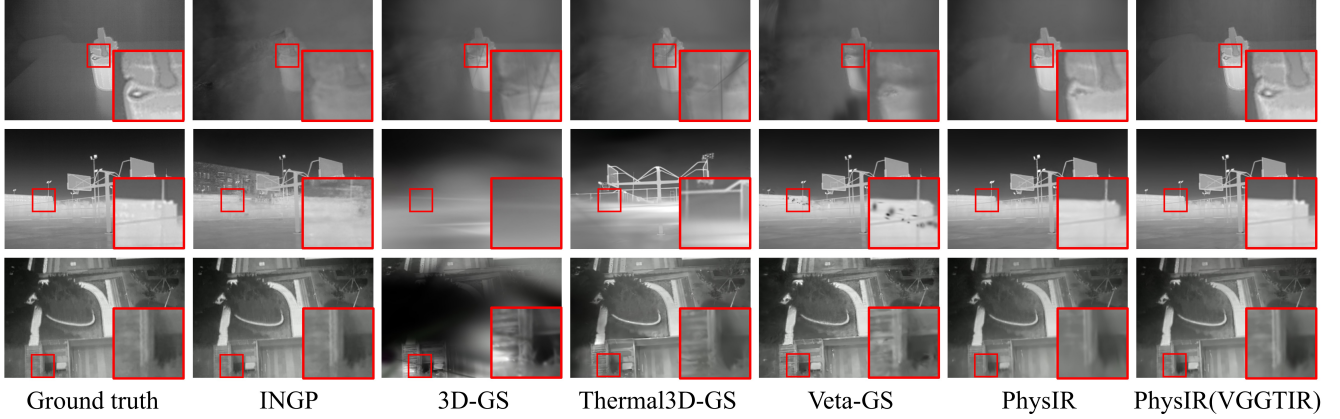


Figure 5. Qualitative comparison against current *state-of-the-art* IR-only methods on the TI-NSD dataset [7].

datasets provide initial point clouds and camera poses estimated by COLMAP, which are used to initialize all compared methods. In addition, our method includes a variant initialized with VGGT-IR.

4.2.1. Comparison with IR-only Methods

Baselines. We evaluate two representative TIR-only reconstruction methods: Veta-GS [35], which models thermal variations using view-dependent deformation fields, and Thermal3D-GS [7], which augments 3DGS rendering with physically motivated mechanisms; 3DGS [25] and InstantNGP [34] are included as references.

Dataset. We evaluate scenes ranging from small to large scale (resolution 480×720) on TI-NSD [7], which were captured with a Zenmuse XT thermal camera spanning a spectral range of $7.5\text{--}13.5\ \mu\text{m}$.

Results. As shown in Tab. 2, **PhysIR-Splat+VGGT-IR** achieves the best performance on Indoor/Outdoor/UAV and on the overall average: mean PSNR improves from 35.97 to 37.33 dB (+1.36 dB), SSIM from 0.958 to 0.962, and LPIPS drops from 0.169 to 0.150. Fig. 5 further shows: InstantNGP is blurrier in detail; 3DGS and Thermal3D-GS exhibit local Gaussian misalignment, causing geometry-texture inconsistency; Veta-GS shows “floating” artifacts under strong deformation or large parallax. In contrast,

PhysIR-Splat preserves edges and thermal gradient details better; combined with **VGGT-IR**, geometry and thermal radiation modeling are more consistent, with fewer artifacts.

4.2.2. Comparison with IR+RGB Multimodal Methods

Baselines. We compare with ThermoNeRF [18], which guides density using RGB while suppressing cross-modal leakage, and ThermalGaussian [30], which adopts multimodal initialization with thermal constraints; 3DGS is included as a reference.

Dataset. We evaluate on RGBT-Scenes [30]. Thermal infrared data were captured with a FLIR E6 Pro (spectral range $7.5\ \mu\text{m}$ to $13\ \mu\text{m}$). RGBT-Scenes provides Ironbow pseudo-colored TIR. We perform a one-off per-scene calibration: sample the color bar to build a color→temperature (linear grayscale) lookup, and convert pseudo-color to a single-channel thermal signal per pixel. Grayscale is used for training; for evaluation/visualization we forward-colorize using the same color bar.

Results. As shown in Table 3, our method significantly outperforms all baselines on both THERMAL and RGB modalities across PSNR, SSIM, and LPIPS. The IR-specific opacity facilitates cross-modality decoupling, further improving overall reconstruction quality. **PhysIR-Splat+VGGT-IR** attains weighted-average

Table 3. Quantitative comparison with state-of-the-art (IR+RGB) methods on the RGBT-Scenes dataset [30].

Metric	Method	THERMAL					RGB				
		Dimsum	DailyStuff	Ebike	Building	Avg.	Dimsum	DailyStuff	Ebike	Building	Avg.
PSNR $\uparrow$	3DGS [25]	26.35	18.77	20.89	25.76	24.52	23.91	20.43	26.77	20.95	22.20
	ThermoNeRF [18]	24.27	17.34	19.70	23.31	22.42	19.74	16.79	17.75	17.12	17.86
	ThermalGaussian [30]	26.46	22.28	23.31	26.72	25.73	24.38	21.76	26.85	24.19	24.13
	Ours	<u>27.61</u>	<u>24.89</u>	<u>25.48</u>	<u>27.82</u>	<u>27.15</u>	<u>25.32</u>	<u>23.76</u>	<u>27.39</u>	<u>25.01</u>	<u>25.12</u>
	Ours(VGGT-IR)	29.41	27.74	27.89	29.50	29.09	27.72	26.02	29.28	27.37	27.44
SSIM $\uparrow$	3DGS [25]	0.889	0.789	0.806	0.872	0.859	0.847	0.748	0.901	0.791	0.810
	ThermoNeRF [18]	0.747	0.759	0.694	0.804	0.772	0.688	0.639	0.540	0.460	0.556
	ThermalGaussian [30]	0.886	0.835	0.862	0.888	0.878	0.858	0.797	0.905	0.849	0.849
	Ours	<u>0.897</u>	<u>0.852</u>	<u>0.878</u>	<u>0.896</u>	<u>0.889</u>	<u>0.876</u>	<u>0.829</u>	<u>0.914</u>	<u>0.866</u>	<u>0.868</u>
	Ours(VGGT-IR)	0.909	0.863	0.889	0.912	0.902	0.891	0.859	0.929	0.881	0.885
LPIPS $\downarrow$	3DGS [25]	0.124	0.274	0.313	0.211	0.204	0.194	0.299	0.171	0.228	0.224
	ThermoNeRF [18]	0.312	0.494	0.290	0.234	0.297	0.228	0.465	0.244	0.291	0.294
	ThermalGaussian [30]	0.129	0.210	0.203	0.177	0.171	0.194	0.253	0.169	0.170	0.188
	Ours	<u>0.106</u>	<u>0.178</u>	<u>0.174</u>	<u>0.150</u>	<u>0.144</u>	<u>0.185</u>	<u>0.223</u>	<u>0.143</u>	<u>0.142</u>	<u>0.166</u>
	Ours(VGGT-IR)	0.082	0.159	0.152	0.122	0.119	0.161	0.201	0.115	0.120	0.142

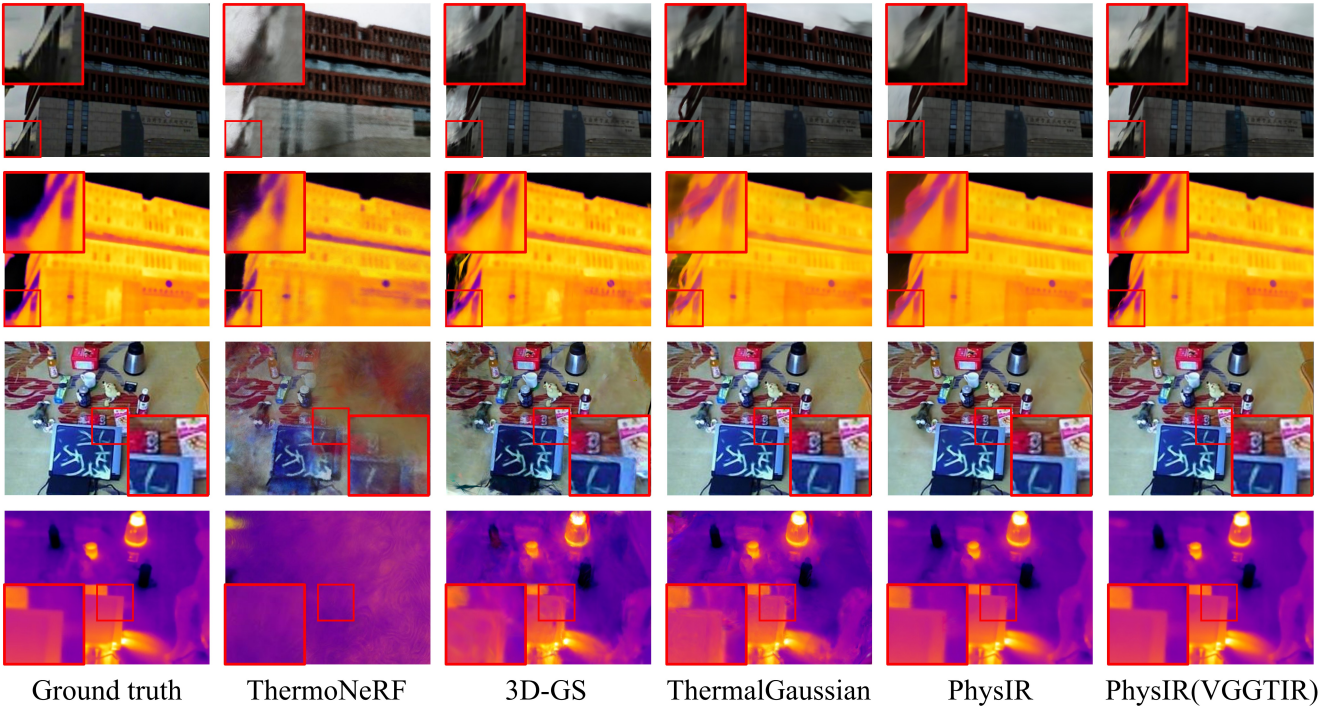


Figure 6. Qualitative comparison against current *state-of-the-art* IR+RGB methods on the RGBT-Scenes [30] (Building, DailyStuff).

PSNRs of 29.09/27.44 on THERMAL/RGB, SSIMs of 0.902/0.885, and reduces LPIPS to 0.119/0.142. Qualitative comparisons in Fig. 6 show that ThermoNeRF tends to produce overly smoothed outputs with blurred details, and occasionally suffers large-scale reconstruction collapse in low-texture regions or under cross-view radiometric inconsistency; 3DGS and ThermalGaussian frequently exhibit

prominent floating artifacts and background leakage. In contrast, our approach yields sharper boundaries and more consistent thermal edges while substantially suppressing artifacts.

4.2.3. Emissivity correlation analysis

On RGB views of RGBT-Scenes, we select several 5×5 pixel patches. Material categories and their reference emis-

Table 4. Ablation study on the TI-NSD datasets [7]. “Atm. Atten.” stands for “Atmospheric Attenuation.”

Method	Indoor			Outdoor			UAV			Average		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Full	38.41	0.967	0.228	35.57	0.952	0.126	38.12	0.969	0.096	37.33	0.962	0.150
w/o VGGT-IR	37.83	0.964	0.236	34.25	0.950	0.138	37.46	0.967	0.098	36.47	0.960	0.158
w/o IR Opacity	37.20	0.963	0.245	34.76	0.948	0.140	37.53	0.965	0.102	36.46	0.958	0.162
w/o Atm. Atten.	38.11	0.965	0.232	34.50	0.945	0.135	37.20	0.962	0.105	36.56	0.957	0.157

sivities in the LWIR band (7.5–13 μm) are taken from public sources and literature [6, 12, 13, 23, 36, 44]. For each location, we export from the thermal image renderer the forward compositing path along the camera ray (top 16 Gaussians and their contribution weights). Using these weights, we perform a weighted read-out of the model emissivity and average within the patch. Figure. 7 shows that most common materials cluster near the diagonal, with a few (e.g., glass and steel) deviating. These results indicate that our method provides broadly consistent and physically reasonable emissivity estimates for the vast majority of common materials in the scene; for the small set of highly challenging materials (e.g., glass/metal with strong specular characteristics), the observed deviations are consistent with their material properties and the inherent limitations of thermal infrared (TIR) imaging.

4.3. Ablation Study

We ablate three components: (i) VGGT-IR for pose and initial geometry, (ii) an IR-specific opacity, and (iii) atmospheric transmittance. Experiments on TI-NSD show that removing any single component causes a clear drop (Table 4), especially in long-range, low-texture scenes and under stronger cross-view radiometric inconsistency, indicating their complementary roles in enforcing physical consistency and stabilizing optimization.

VGGT-IR. Replacing VGGT-IR with COLMAP reduces the averages by PSNR $\downarrow$ 2.3%, SSIM $\downarrow$ 0.2%, LPIPS $\uparrow$ 5.3%, with a larger outdoor drop (PSNR $\downarrow$ 3.7%). This confirms that VGGT-IR is better matched to TIR sequences and yields more reliable initialization.

IR Opacity. Employing an IR-specific opacity α_i^{IR} (Eq. 11) better captures the absorption–self-emission mechanism inherent to TIR imaging. Removing it (reverting to a shared opacity [30]) degrades averages by PSNR $\downarrow$ 2.3%, SSIM $\downarrow$ 0.4%, LPIPS $\uparrow$ 8.0%, with greater impact in Indoor scenes (PSNR $\downarrow$ 3.1%), validating that IR-tailored opacity decoupling effectively mitigates cross-modality entanglement.

Atmospheric Transmittance. Incorporating the distance-dependent $T_{\text{atm}}(d_i)$ (Eq. 12) is advantageous for modeling atmospheric effects in long-range TIR captures:

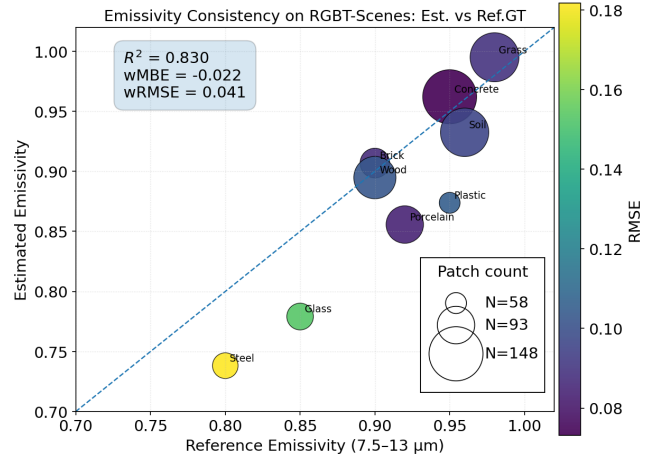


Figure 7. Horizontal axis: reference emissivity from literature (Lit.Ref., LWIR 7.5–13 μm); vertical axis: model estimate (Est.). Each bubble corresponds to one material; bubble size $\propto$ number of sample patches (N); color encodes per-material RMSE. The dashed line ($y = x$) indicates ideal agreement.

omitting it diminishes averages by PSNR $\downarrow$ 2.1%, SSIM $\downarrow$ 0.5%, LPIPS $\uparrow$ 4.7%, with a pronounced Outdoor decline (PSNR $\downarrow$ 3.0%). This supports modeling transmittance to prevent overestimation of far-field radiance and the associated biases.

5. Conclusion

This paper departs from the conventional practice of treating thermal infrared (TIR) radiance as color channels, and instead explicitly models temperature and emissivity on Gaussian primitives while integrating TIR radiative transfer into the 3DGS rendering pipeline. In addition, we introduce VGGT-IR, a feed-forward initializer that recovers scene geometry and camera poses directly from TIR sequences, demonstrating superior performance across multiple datasets. However, the diffuse-reflection assumption limits the ability to model specular materials (e.g., glass and polished steel); as shown in Fig. 7, these cases exhibit substantial errors, underscoring inherent limitations of the current model.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFA1008600; in part by the National Natural Science Foundation of China under Grant 92470108, and Grant 62272363; in part by the Joint Laboratory for Innovation in Satellite-Borne Computers and Electronics Technology Open Fund 2023 under Grant 2024KFKT001-1.

References

- [1] Manuel López Antequera, Pau Gargallo, Markus Hofinger, Samuel Rota Buló, Yubin Kuang, and Peter Kotschieder. Mapillary planet-scale depth dataset. In *European Conference on Computer Vision*, pages 589–604. Springer, 2020. 5
- [2] Fanglin Bao, Xueji Wang, Shree Hari Sureshbabu, Gautam Sreekumar, Liping Yang, Vaneet Aggarwal, Vishnu N Boddeti, and Zubin Jacob. Heat-assisted detection and ranging. *Nature*, 619(7971):743–748, 2023. 2
- [3] Elizabeth Bondi, Debadepta Dey, Ashish Kapoor, Jim Davis, Shital Shah, Fei Fang, Bistra Dilkina, Robert Hanaford, Arvind Iyer, Lucas Joppa, et al. Airsim-w: A simulation environment for wildlife conservation with uavs. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, pages 1–12, 2018. 1
- [4] Dorit Borrmann, Andreas Nüchter, Marija akulović, Ivan Maurović, Ivan Petrović, Dinko Osmanković, and Jasmin Velagić. A mobile robot based system for fully automated thermal 3d mapping. *Advanced Engineering Informatics*, 28(4):425–440, 2014. 2
- [5] Ivo Campione, Francesca Lucchi, Nicola Santopuoli, and Leonardo Seccia. 3d thermal imaging system with decoupled acquisition for industrial and cultural heritage applications. *Applied Sciences*, 10(3):828, 2020. 2
- [6] Chiachung Chen. Determining the leaf emissivity of three crops by infrared thermometry. *Sensors*, 15(5):11387–11401, 2015. 8
- [7] 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. 1, 2, 6, 8
- [8] Jorge Corral-Acero, Francesca Margara, Maciej Marciniak, Cristobal Rodero, Filip Loncaric, Yingjing Feng, Andrew Gilbert, Joao F Fernandes, Hassaan A Bukhari, Ali Wajdan, et al. The ‘digital twin’ to enable the vision of precision cardiology. *European heart journal*, 41(48):4556–4564, 2020. 1
- [9] Weichen Dai, Yu Zhang, Shenzhou Chen, Donglei Sun, and Da Kong. A multi-spectral dataset for evaluating motion estimation systems. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5560–5566. IEEE, 2021. 5
- [10] Julio Manuel De Luis-Ruiz, Javier Sedano-Cibrián, Rubén Pérez-Álvarez, Raúl Pereda-García, and Ramiro Benito Salas-Menocal. Generation of 3d thermal models for the analysis of energy efficiency in buildings. In *International conference on The Digital Transformation in the Graphic Engineering*, pages 741–754. Springer, 2022. 2
- [11] Yasser El Masri and Tarek Rakha. A scoping review of non-destructive testing (ndt) techniques in building performance diagnostic inspections. *Construction and Building Materials*, 265:120542, 2020. 1
- [12] Fluke Process Instruments. Emissivity - metals. Available online: <https://www.flukeprocessinstruments.com/en-us/service-and-support/knowledge-center/infrared-technology/emissivity-metals>, . Accessed 2025-11-04. 8
- [13] Fluke Process Instruments. Emissivity values - non-metals. Available online: <https://www.flukeprocessinstruments.com/en-us/why-use-pyrometers%3Femissivity-values-non-metals>, . Accessed 2025-11-04. 8
- [14] Alan Gillespie, Shuichi Rokugawa, Tsuneo Matsunaga, J Steven Cothren, Simon Hook, and Anne B Kahle. A temperature and emissivity separation algorithm for advanced spaceborne thermal emission and reflection radiometer (aster) images. *IEEE transactions on geoscience and remote sensing*, 36(4):1113–1126, 1998. 2
- [15] Adam Glowacz. Fault diagnosis of electric impact drills using thermal imaging. *Measurement*, 171:108815, 2021. 1
- [16] Youngjib Ham and Mani Golparvar-Fard. An automated vision-based method for rapid 3d energy performance modeling of existing buildings using thermal and digital imagery. *Advanced Engineering Informatics*, 27(3):395–409, 2013. 2
- [17] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. 2003. 4
- [18] Mariam Hassan, Florent Forest, Olga Fink, and Malcolm Mielle. Thermonerf: A multimodal neural radiance field for joint rgb-thermal novel view synthesis of building facades. *Advanced Engineering Informatics*, 65:103345, 2025. 2, 6, 7
- [19] Ludwig Hoegner and Uwe Stilla. Thermal leakage detection on building facades using infrared textures generated by mobile mapping. In *2009 Joint Urban Remote Sensing Event*, pages 1–6. IEEE, 2009. 2
- [20] L Hoegner and U Stilla. Mobile thermal mapping for matching of infrared images with 3d building models and 3d point clouds. *Quantitative Infrared thermography journal*, 15(2): 252–270, 2018. 2
- [21] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pages 492–518. Springer, 1992. 5
- [22] Glynn Hulley, Robert Freepartner, Nabin Malakar, and Sudipta Sarkar. Moderate resolution imaging spectroradiometer (modis) land surface temperature and emissivity product (mxd21) user guide. NASA: Washington, DC, USA, 26, 2016. 3
- [23] KS Humes, WP Kustas, MS Moran, WD Nichols, and MA Weltz. Variability of emissivity and surface temperature over a sparsely vegetated surface. *Water resources research*, 30(5):1299–1310, 1994. 8
- [24] Dorota Iwaszczuk and Uwe Stilla. Camera pose refinement by matching uncertain 3d building models with thermal infrared image sequences for high quality texture extrac-

- tion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132:33–47, 2017. 2
- [25] 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, 2, 6, 7
- [26] S Lagüela, J Martínez, J Armesto, and P Arias. Energy efficiency studies through 3d laser scanning and thermographic technologies. *Energy and buildings*, 43(6):1216–1221, 2011. 2
- [27] Barid Baran Lahiri, Subramaniam Bagavathiappan, T Jayakumar, and John Philip. Medical applications of infrared thermography: a review. *Infrared physics & technology*, 55(4):221–235, 2012. 1
- [28] Huaizhou Li, Shuaijun Wang, Zhenpeng Bai, Hong Wang, Sen Li, and Shupeí Wen. Research on 3d reconstruction of binocular vision based on thermal infrared. *Sensors*, 23(17):7372, 2023. 1
- [29] Ni Li, Wenqing Huai, Shaodan Wang, and Lei Ren. A real-time infrared imaging simulation method with physical effects modeling of infrared sensors. *Infrared Physics & Technology*, 78:45–57, 2016. 1
- [30] Rongfeng Lu, Hangyu Chen, Zunjie Zhu, Yuhang Qin, Ming Lu, Chenggang Yan, et al. Thermalgaussian: Thermal 3d gaussian splatting. In *The Thirteenth International Conference on Learning Representations*. 1, 2, 6, 7, 8
- [31] Fangyuan Mao, Jilin Mei, Shun Lu, Fuyang Liu, Liang Chen, Fangzhou Zhao, and Yu Hu. Pid: physics-informed diffusion model for infrared image generation. *Pattern Recognition*, 169:111816, 2026. 3, 5
- [32] ELEONORA Maset, Andrea Fusiello, Fabio Crosilla, R Toldo, D Zorzetto, et al. Photogrammetric 3d building reconstruction from thermal images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:25–32, 2017. 2
- [33] 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. 2
- [34] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 6
- [35] Myeongseok Nam, Wongi Park, Minsol Kim, Hyejin Hur, and Soomok Lee. Veta-gs: View-dependent deformable 3d gaussian splatting for thermal infrared novel-view synthesis. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 965–970. IEEE, 2025. 1, 2, 6
- [36] Omega Engineering. Table of total emissivity. Available online: <https://assets.omega.com/tech-ref-legacy-pdfs/z08-8-089.pdf>. Accessed 2025-11-04. 8
- [37] Jose Pedro Aguerre and Eduardo Fernández. Thrend: A ray tracing module for infrared rendering of urban scenes. In *Building Simulation 2021*, pages 1588–1595. IBPSA, 2021. 1
- [38] Ravi Ramamoorthi and Pat Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 497–500, 2001. 3
- [39] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 12179–12188, 2021. 4
- [40] Johannes Rangel, Samuel Soldan, and A Kroll. 3d thermal imaging: Fusion of thermography and depth cameras. In *International Conference on Quantitative InfraRed Thermography*, 2014. 1
- [41] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 2, 5
- [42] Sebastian Schramm, Phil Osterhold, Robert Schmoll, and Andreas Kroll. Combining modern 3d reconstruction and thermal imaging: Generation of large-scale 3d thermograms in real-time. *Quantitative InfraRed Thermography Journal*, 19(5):295–311, 2022. 2
- [43] Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025. 2, 4
- [44] ThermoWorks. Infrared emissivity table. Available online: <https://www.thermoworks.com/emissivity-table/>. Accessed 2025-11-04. 8
- [45] Stephen Vidas and Peyman Moghadam. Heatwave: A handheld 3d thermography system for energy auditing. *Energy and buildings*, 66:445–460, 2013. 2
- [46] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5294–5306, 2025. 2, 3, 4, 5
- [47] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021. 4
- [48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5
- [49] Meng Wei, Qianyi Wu, Jianmin Zheng, Hamid Reza Tofighi, and Jianfei Cai. Normal-gs: 3d gaussian splatting with normal-involved rendering. *Advances in Neural Information Processing Systems*, 37:76356–76379, 2024. 4
- [50] Xin Wu, Jianqi Zhang, Yang Chen, and Xi Huang. Real-time mid-wavelength infrared scene rendering with a feasible brdf model. *Infrared Physics & Technology*, 68:124–133, 2015. 1
- [51] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1790–1799, 2020. 5
- [52] Tianxiang Ye, Qi Wu, Junyuan Deng, Guoqing Liu, Liu Liu, Songpengcheng Xia, Liang Pang, Wenxian Yu, and Ling Pei.

- Thermal-nerf: Neural radiance fields from an infrared camera. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1046–1053. IEEE, 2024. [2](#)
- [53] Chandan Yeshwanth, Yueh-Cheng Liu, Matthias Nießner, and Angela Dai. Scannet++: A high-fidelity dataset of 3d indoor scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12–22, 2023. [5](#)
- [54] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. [5](#)
- [55] Shibo Zhao, Zheng Fang, and Shiguang Wen. A real-time handheld 3d temperature field reconstruction system. In *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 289–294. IEEE, 2017. [1](#)
- [56] Zheng Zhou, Yaqoob Majeed, Geraldine Diverres Naranjo, and Elena MT Gambacorta. Assessment for crop water stress with infrared thermal imagery in precision agriculture: A review and future prospects for deep learning applications. *Computers and Electronics in Agriculture*, 182: 106019, 2021. [1](#)
- [57] Jingwei Zhu, Yusheng Xu, Zhen Ye, Ludwig Hoegner, and Uwe Stilla. Fusion of urban 3d point clouds with thermal attributes using mls data and tir image sequences. *Infrared Physics & Technology*, 113:103622, 2021. [2](#)
- [58] Jingwei Zhu, Yusheng Xu, Ludwig Hoegner, and Uwe Stilla. Generation of thermal point clouds from uncalibrated thermal infrared image sequences and mobile laser scans. *IEEE Transactions on Instrumentation and Measurement*, 72:1–16, 2023. [2](#)