S³-TIR: Coupled Spatial-Spectral Splatting for Thermal Infrared Novel View Synthesis

Yaoxing Wang[1], Yuheng Li[1], Yan Di[2], Shan Gao[1]*

[1]Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{wangyx24, yuhengli}@mail.nwpu.edu.cn, diyan@hit.edu.cn, gaoshan@nwpu.edu.cn

Abstract

We present **S³-TIR** (**C**oupled **S**patial-**S**pectral **S**platting), a physics-inspired framework for high-fidelity thermal infrared (TIR) novel view synthesis. Existing 3D Gaussian splatting methods struggle to capture the unique duality of TIR imagery: the coexistence of diffusive smoothness from heat conduction and structural sharpness from material discontinuities. To address this issue, we leverage the intrinsic frequency properties of thermal scenes, where conduction creates smooth fields while discontinuities cause sharp signal jumps. Accordingly, S^3-TIR couples adaptive spatial support with internal spectral modulation via a unified generalized-exponential Gabor primitive. This design utilizes spatial support as a physical truncation controller, confining high-frequency textures within sharp boundaries while maintaining smooth diffusion elsewhere. Furthermore, we propose a frequency-aware deferred splatting scheme that enforces multi-view frequency consistency, anchoring frequency distributions to the 3D structure rather than view-dependent appearance. Experiments on diverse benchmarks demonstrate that S^3-TIR achieves state-of-the-art performance, notably reducing LPIPS by 42.5% and improving PSNR by 3.2% on the RGBT-Scenes dataset, while yielding structurally stable thermal sequences free from jittering artifacts.

1 Introduction

High-fidelity Novel View Synthesis is a fundamental problem in computer vision and graphics that enables applications ranging from digital twins to autonomous navigation. Given the recent progress of explicit neural representations, 3D Gaussian Splatting (3DGS) [Kerbl *et al.*, 2023] has become a dominant standard for synthesizing photorealistic RGB scenes in real-time. However, real-world sensing goes beyond visible light. Thermal Infrared (TIR) imaging offers unique capabilities for safety-critical scenarios[Shi *et al.*, 2023; Shi *et al.*, 2024; Hu *et al.*, 2025; Bao *et al.*, 2025], ranging from drone-based fire detection [Wang *et al.*, 2025b; Chen *et al.*, 2004; Lin *et al.*, 2all; Lin *et al.*, 2026] to medical diagnostics [Ring and Ammer, 2012; Ma *et al.*, 2023], by characterizing thermodynamic states in total darkness and penetrating atmospheric obscurant [Wang *et al.*, 2023].

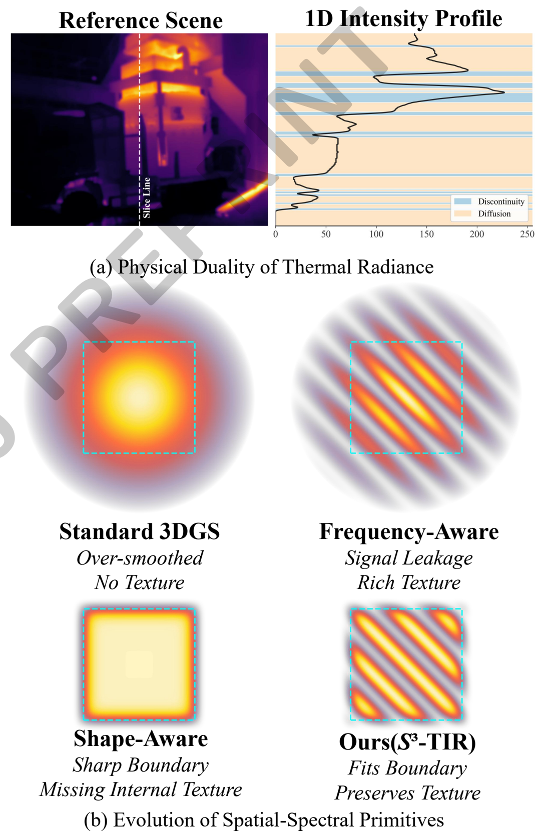

Figure 1: **Structural Duality of Thermal Radiance and Primitive Evolution. (a)** The 1D intensity profile reveals the intrinsic conflict in thermal scenes: the coexistence of *diffusive smoothness* (orange) driven by heat conduction and *structural discontinuities* (blue). **(b)** Comparison of primitives. Standard 3DGS blurs edges due to its low-pass nature. Frequency-aware methods capture texture but suffer from *signal leakage*. Shape-aware methods fit boundaries but miss internal content. In contrast, our **S³-TIR** explicitly couples spatial support with spectral modulation, enabling *physical truncation* of textures at sharp boundaries.

*Corresponding author.

A straightforward idea is to directly extend standard 3DGS to thermal data. However, such methods often fail to produce structurally stable results [Song *et al.*, 2024; Kheradmand *et al.*, 2024; Liu *et al.*, 2025a]. Smooth thermal fields may appear noisy, sharp material edges may blur, and high-frequency details often exhibit severe jitter artifacts in novel views. The key reason behind this limitation lies in the intricate nature of thermal radiation. Treating thermal images as simple color textures without capturing the underlying physics leads to a fundamental representation mismatch.

This paper aims to explicitly model the 3D distribution of thermal signals by analyzing the physical mechanisms underlying their appearance. Our approach is premised on the fundamental observation of the structural duality of thermal radiance (see Fig. 1(a)), wherein the frequency characteristics of a 3D point constitute an intrinsic, view-independent physical property governed by two opposing processes. First, **heat conduction** [Hahn and Özisik, 2012; Widder, 1976; Bao *et al.*, 2023], governed by Fourier's law [Narasimhan, 1999; Hahn and Özisik, 2012], drives thermal energy to diffuse continuously across object surfaces. This process mathematically approximates a low-pass filter, yielding a dominant diffusive smoothness. Second, **Material Discontinuities**, such as abrupt variations in emissivity or geometric occlusions, manifest as extremely sharp, step-like signal jumps [Widder, 1976]. These jumps introduce structural sharpness and sparse high-frequency details. However, reconciling these conflicting attributes poses a significant challenge due to decomposition ambiguity, where a sharp thermal edge can be mathematically interpreted as either a spatially sharp envelope or a high-frequency internal texture. Standard direct rendering masks this distinction, leading the network to overfit view-dependent artifacts rather than inferring the true physical structure. By addressing these factors, we bridge the gap between thermal physics and 3D representation within a unified framework.

Building upon this insight, we introduce $\mathbf{S^3}$**-TIR**, a physics-inspired framework designed to reconcile the coexistence of diffusive smoothness and structural sharpness. To address this, we first propose a unified generalized-exponential gabor primitive. By explicitly coupling a generalized-exponential spatial support with an internal spectral modulation field, we introduce a learnable shape parameter β acting as a physical truncation controller. This mechanism enables the primitive to adaptively transition from diffusive gaussian falloffs to sharp box-like boundaries, strictly confining high-frequency textures within the geometric envelope to prevent signal leakage. Crucially, to ensure a physically plausible decomposition, we implement a frequency-aware deferred splatting pipeline. Instead of directly rendering the final color, we independently rasterize intrinsic physical attributes into observable frequency g-buffers, including the low-frequency energy, high-frequency energy, and spatial sharpness maps. Leveraging these intermediate representations, we enforce a multi-view frequency consistency loss that anchors the frequency distribution to the 3d geometry across perspectives. This effectively resolves the decomposition ambiguity and prevents high-frequency jittering. Quantitative and qualitative results on benchmarks like rgbt-scenes and tinsd demonstrate that $\mathbf{S^3}$-TIR achieves state-of-the-art performance, notably reducing lpips by 42.5% and improving psnr by 3.2% on the challenging rgbt-scenes dataset, while yielding structurally stable thermal sequences free from physically implausible artifacts.

In summary, our main contributions are as follows:

1. **Coupled Spatial-Spectral Representation:** We propose a unified generalized-exponential Gabor primitive that physically truncates internal spectral modulation at geometric boundaries, reconciling diffusive smoothness with structural sharpness.
2. **Frequency-Aware Ambiguity Resolution:** We introduce a deferred splatting pipeline with a multi-view frequency consistency loss, which anchors rasterized physical attributes to the 3D structure and resolves decomposition ambiguity.
3. **Superior Performance:** $\mathbf{S^3}$-TIR achieves state-of-the-art performance on multiple benchmarks, setting a new standard for structurally stable thermal synthesis free from jittering artifacts.

2 Related Work

2.1 Physics-Informed Thermal Rendering

Early thermal 3D reconstruction heavily relied on fusing thermal imagery with auxiliary depth sensors or RGB-D systems to retrieve geometric clues from textureless regions [Rangel *et al.*, 2014; Vidas and Moghadam, 2013; Zhao *et al.*, 2017]. With the paradigm shift to neural rendering, implicit representations first adapted the NeRF framework to the thermal domain to enable continuous scene representation [Hassan *et al.*, 2024; Ye *et al.*, 2024], followed by recent explicit 3D Gaussian Splatting (3DGS) approaches that incorporate constraints for atmospheric transmission [Chen *et al.*, 2024; Lu *et al.*,] or dynamic temporal evolution [Yang *et al.*, 2025; Liu *et al.*, 2025b; Liu *et al.*, 2024]. Beyond synthesis, emerging research aims to ground vision models in physical laws, utilizing heat conduction [Wang *et al.*, 2025c] or thermodynamic modeling [Chen *et al.*, 2025; Wang *et al.*, 2026] to enhance consistency. However, unlike visible spectrum methods that leverage physics for inverse rendering [Liang *et al.*, 2024; Gao *et al.*, 2024; Jiang *et al.*, 2024], explicitly inverting intrinsic thermal parameters remains a highly ill-posed problem, especially when accurate ground-truth supervision is unavailable. This limitation motivates our approach to focus on developing a more expressive representation capable of capturing the unique duality of thermal radiation, which is defined by both diffusive smoothness and structural sharpness, without relying on intractable inverse physics solving.

2.2 Frequency-Aware and Shape-Adaptive 3DGS

To overcome the inherent low-pass nature of standard 3DGS [Kerbl *et al.*, 2023], two distinct research directions have emerged. **Frequency-aware methods** introduce Fourier regularization [Zhang *et al.*, 2024a; Zeng *et al.*, 2025; Ren *et al.*, 2025] or replace Gaussians with Wavelet and Gabor-modulated primitives [Zhang *et al.*, 2025a; Zhao *et al.*, 2025a; Zhou *et al.*, 2025; Wurster *et al.*, 2024; Zhang *et al.*, 2025b] to capture high-frequency textures. However, these methods

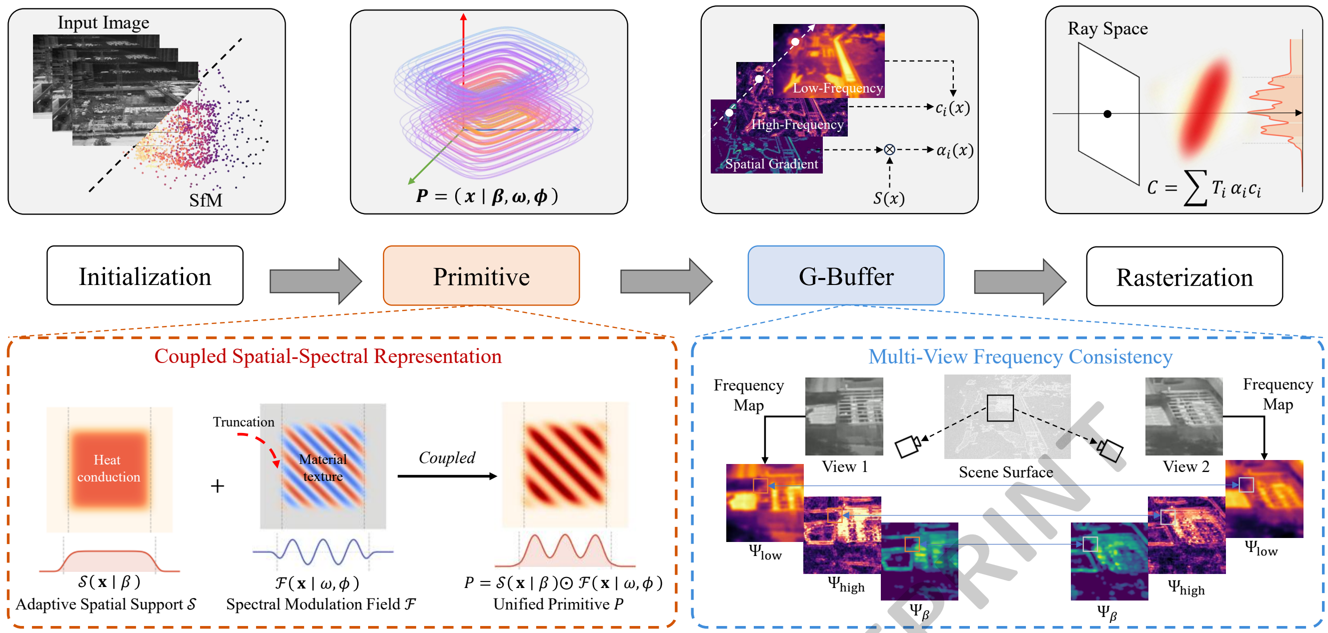

Figure 2: **Overall pipeline of S^3-TIR.** A physics-inspired framework initializes Unified Primitives from SfM points. Through the *Coupled Spatial-Spectral Representation*, the model optimizes an adaptive spatial support $\mathcal{S}$ and predicts internal spectral modulation $\mathcal{F}$ to capture heat conduction and material texture. The *Frequency-Aware Deferred Splatting* decouples the thermal signal into intermediate G-Buffers ($\Psi_{low}, \Psi_{high}, \Psi_{\beta}$), i.e., intrinsic physical attributes are rasterized separately rather than as a final color. physics-inspired constraints, specifically the *Multi-View Frequency Consistency*, are enforced to anchor these properties in 3D space, ensuring that frequency distributions projected from different views (View 1, View 2) remain consistent on the scene surface.

typically treat frequency as a texture attribute independent of geometry, lacking the mechanism to spatially truncate signals at sharp boundaries, which often results in high-frequency energy spilling into empty space. Conversely, **shape-adaptive approaches** employ generalized-exponential falloffs [Hamdi *et al.*, 2024] or learnable deformations and ensembling strategies [Di *et al.*, 2024; Zhang *et al.*, 2024b; Li *et al.*, 2025; Huang *et al.*, 2025; Zhao *et al.*, 2025b] to fit sharp geometric silhouettes. Other works enforce strict geometric constraints by flattening Gaussians into 2D disks [Huang *et al.*, 2024; Yu *et al.*, 2024]. While effective for boundaries and surfaces, these methods oversimplify internal content (e.g., assuming homochromatic interiors) and fail to capture complex thermal gradients driven by heat diffusion [Wang *et al.*, 2025a], rendering them insufficient for physically consistent thermal modeling. Our S^3-TIR bridges this gap by explicitly coupling adaptive spatial truncation with internal spectral modulation, thereby reconciling the conflict between geometric sharpness and diffusive smoothness within a unified primitive.

3 Method

We present **S^3**-TIR, a physics-inspired framework designed to reconstruct high-fidelity thermal radiance fields from posed TIR images by explicitly reconciling the physical coexistence of diffusive smoothness and structural sharpness. We represent the scene using a set of unified primitives $\mathcal{P} = \{P_n\}$, where each primitive $P_n = (\mathcal{G}_n, \mathcal{F}_n)$ is decomposed into a spatial geometry state $\mathcal{G}_n$ and a spectral content state $\mathcal{F}_n$.

Specifically, the geometry state $\mathcal{G}_n = \{\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n, \beta_n\}$ defines the spatial support via a Generalized-Exponential formulation, where $\boldsymbol{\mu}_n$ denotes the 3D center position, $\boldsymbol{\Sigma}_n$ represents the anisotropic covariance matrix determining scale and rotation, and β_n serves as a shape parameter to control the steepness of boundary sharpness. Complementarily, the spectral state $\mathcal{F}_n = \{\omega_{\text{low},n}, \omega_{\text{high},n}, \mathbf{f}_n, \phi_n\}$ parameterizes the internal thermal distribution via Gabor modulation, consisting of a base thermal intensity $\omega_{\text{low},n}$, a modulation amplitude $\omega_{\text{high},n}$, a spatial frequency vector $\mathbf{f}_n$, and a phase shift ϕ_n. The complete optimization pipeline is illustrated in Fig. 2.

3.1 Coupled Spatial-Spectral Primitive

The Gap in Thermal Representation. Thermal scenes inherently combine macroscopic geometry with fine-scale textures. Standard 3DGS, limited by its Gaussian low-pass nature, models heat diffusion well but struggles to capture material discontinuities. While recent variants like GES or Gabor wavelets improve shape or texture respectively, they treat spatial geometry and frequency content in isolation. Specifically, shape-focused methods lack the spectral capacity for textures, while frequency-focused methods fail to truncate signals at sharp boundaries. To bridge this, we propose a unified primitive that models the thermal field as a *spatially-constrained spectral signal*, enforcing a direct coupling between geometric sharpness and frequency distribution.

Primitive Formulation. We represent the scene using a set of primitives $\mathcal{P} = \{P_i\}_{i=1}^N$. Departing from traditional methods that assume a fixed Gaussian falloff, we formulate each primitive $P(\mathbf{x})$ as the coupled product of an *Adaptive Spatial Support* $\mathcal{S}$ and a *Spectral Modulation Field* $\mathcal{F}$:

$$\mathcal{P}(\mathbf{x}) = \mathcal{S}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \beta) \cdot \mathcal{F}(\mathbf{x} \mid \omega, \mathbf{f}, \phi) \tag{1}$$

1) Adaptive Spatial Support ($\mathcal{S}$). This component defines the effective physical domain of the primitive. We employ a Generalized-Exponential formulation to govern the spatial falloff:

$$\mathcal{S}(\mathbf{x}) = \exp\left(-\left\|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu})\right\|^{\beta}\right) \tag{2}$$

Here, the shape parameter $\beta \in [2, \infty)$ serves as a physical truncation controller:

- **Diffusion Mode ($\beta \approx 2$):** The support approximates a standard Gaussian, effectively approximating smooth thermal diffusion where intensity naturally decays across surfaces.
- **Discontinuity Mode ($\beta \to \infty$):** The support converges to a box-like function. Crucially, this mode spatially truncates the internal spectral content at geometric boundaries, preventing signal leakage. This allows a single primitive to represent a textured step function—a capability absent in fixed-Gaussian methods.

2) Spectral Modulation Field ($\mathcal{F}$). Within the spatial support defined by $\mathcal{S}$, we model intricate thermal variations using a phase-aware frequency formulation:

$$\mathcal{F}(\mathbf{x}) = \mathrm{ReLU}\Bigg(\omega_{\text{low}} + \sum_{j=1}^{M} \omega_{j,\text{high}} \times \cos\left(2\pi \mathbf{f}_j^{\top}(\mathbf{x}-\boldsymbol{\mu}) + \phi_j\right)\Bigg) \tag{3}$$

By explicitly parameterizing the 3D frequency vector $\mathbf{f}_j$, this term captures anisotropic thermal textures (e.g., directional heat dissipation patterns) that cannot be adequately represented by spherical harmonics alone.

3.2 Frequency-Aware Deferred Rendering

The Frequency Decoupling Ambiguity. Existing frequency-aware 3DGS methods typically modulate Gaussian primitives with high-frequency signals, operating under the implicit assumption that geometric boundaries and internal textures are naturally separable. We contend that this assumption breaks down at material discontinuities. Due to the intrinsic smooth falloff of standard primitives with $\beta = 2$, these methods unavoidably suffer from *Signal Leakage* as illustrated in Fig. 3. This introduces a fundamental ambiguity in the solver, making it an ill-posed problem to distinguish between the physically sharp spatial envelope of *Geometric Truncation* and the artificially decaying texture of *Spectral Overfitting*. Consequently, rather than recovering the underlying scene structure, the model compensates for this ambiguity by generating spurious high-frequency artifacts that degrade structural fidelity.

Frequency-Aware G-Buffer Generation. To address this ambiguity, we propose a deferred rendering pipeline that explicitly decouples thermal signals. Drawing inspiration from inverse rendering frameworks that separate material properties (e.g., albedo, normal) via deferred buffers [Liang *et al.*, 2024; Gao *et al.*, 2024], we here repurpose the rasterization engine to independently project the intrinsic spatial and spectral attributes of primitives into observable 2D feature maps, termed *Frequency G-Buffers*. Formally, this process accumulates specific physical properties for each pixel $\mathbf{u}$ via alpha-blending:

1. **Low-Frequency Energy Map (Ψ_{low}):** Isolates the base thermal intensity ω_{low} to recover the smooth diffusive background;
2. **High-Frequency Energy Map (Ψ_{high}):** Extracts the Gabor modulation energy $\sum|\omega_{j,\text{high}}|$ to reveal internal structural details;
3. **Spatial Sharpness Map (Ψ_β):** Projects the shape parameter β to expose geometric discontinuity boundaries.

To efficiently support this multi-channel rasterization while ensuring valid coverage of the Generalized-Exponential support, we employ a *Bounding Gaussian Strategy*, utilizing the standard Gaussian ($\beta = 2$) as a geometric proxy for sorting and tile-culling.

Multi-view Frequency Consistency. The pivotal advantage of this deferred architecture is the ability to constrain the 2D G-Buffers to produce multi-view consistent frequency maps. Drawing inspiration from Multi-View Stereo (MVS) constraints, we enforce that for a surface point visible from both viewpoint v_i and v_j, its projected frequency attributes on maps Ψ_i and Ψ_j must be identical. Specifically, we warp

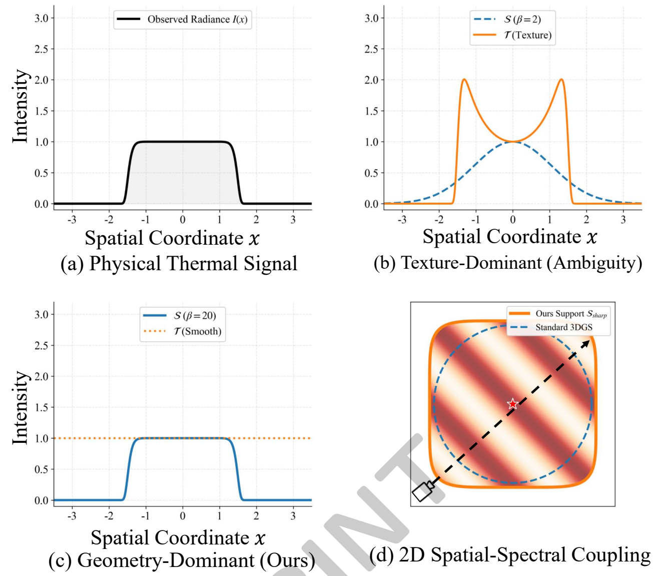

Figure 3: **Mechanism of Spatial-Spectral Coupling.** (a) Physical thermal signal characterized by sharp transitions. (b) Standard 3DGS ($\beta = 2$): the smooth Gaussian fall-off causes signal leakage, leading to texture overfitting. (c) S^3-TIR ($\beta = 20$): the sharp spatial support effectively captures geometry while maintaining smooth and consistent texture. (d) 2D comparison: our $\mathcal{S}_{sharp}$ physically anchors Gabor modulations within geometric boundaries to prevent spectral leakage and reconstruction artifacts.

the G-Buffers from a source view v_j to a reference view v_i using the rendered depth D_i and scene geometry:

$$\Psi_{j\to i}^{\text{warped}}(\mathbf{u}) = \Psi_j\left(\mathbf{K}_j\left(\mathbf{R}_{ij}\mathbf{D}_i(\mathbf{u})\mathbf{K}_i^{-1}\mathbf{u} + \mathbf{T}_{ij}\right)\right) \tag{4}$$

where $\mathbf{K}, \mathbf{R}, \mathbf{T}$ denote intrinsic and extrinsic parameters. We then impose a consistency loss on the physical attribute maps (specifically Ψ_{high} and Ψ_β):

$$\mathcal{L}_{\text{consist}} = \sum_{\Psi\in\{\Psi_{\text{high}},\Psi_\beta\}} \|\Psi_i(\mathbf{u}) - \Psi_{j\to i}^{\text{warped}}(\mathbf{u})\|_1 \tag{5}$$

By *anchoring* frequency properties to the 3D geometry, this constraint eliminates ambiguity, ensuring that structural sharpness and textural details remain geometrically consistent across all perspectives.

3.3 Physics-inspired Hybrid Optimization

To ensure a structurally consistent decomposition of spatial support and spectral content, we employ a hybrid optimization strategy. The total objective function is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{render}} + \lambda_c\mathcal{L}_{\text{consist}} + \lambda_s\mathcal{L}_{\text{sparse}} + \lambda_g\mathcal{L}_{\text{smooth}} \tag{6}$$

where $\mathcal{L}_{\text{render}}$ comprises standard ℓ_1 and D-SSIM losses, and $\mathcal{L}_{\text{consist}}$ enforces the multi-view frequency consistency. To further regularize the intrinsic attributes, we impose two physics-informed priors on the intermediate G-Buffers.

1) Thermodynamic Sparsity ($\mathcal{L}_{\text{sparse}}$). Since heat conduction acts as a natural low-pass filter, high-frequency energy in thermal fields is inherently sparse, concentrating primarily at material discontinuities. We enforce this via an ℓ_1-sparsity penalty on the High-Frequency Energy Map Ψ_{high}:

$$\mathcal{L}_{\text{sparse}} = \frac{1}{HW}\sum_{\mathbf{u}\in\Omega}|\Psi_{\text{high}}(\mathbf{u})| \tag{7}$$

This prior encourages the model to explain the scene using the low-frequency base ω_{low} and adaptive spatial support β, reserving spectral modulation for necessary textural details.

2) Geometric Continuity ($\mathcal{L}_{\text{smooth}}$). Material boundaries, parameterized by the spatial sharpness β, typically exhibit piecewise smoothness rather than random noise. To enforce clean geometric transitions, we apply an isotropic Total Variation (TV) regularizer on the Spatial Sharpness Map Ψ_β:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{HW}\sum_{\mathbf{u}\in\Omega}\sqrt{(\Delta_x\Psi_\beta)^2 + (\Delta_y\Psi_\beta)^2} \tag{8}$$

where Δ denotes spatial gradients. This term suppresses artifacts in the shape parameter while preserving sharp physical discontinuities. Jointly, these constraints force the primitives to learn a structured decomposition by handling smooth regions via diffusion and edges via geometric truncation, thereby eliminating high-frequency jittering.

4 Experiments

4.1 Experimental Setup

Datasets. To ensure robustness, we evaluate on three distinct benchmarks. **RGBT-Scenes** provides geometrically complex handheld data with over $1,000$ aligned pairs. **TI-NSD** offers large-scale diversity ($6,664$ images) across indoor, outdoor, and UAV scenarios with significant thermal variance. Finally, **NTR** serves as a challenging testbed for dynamic environments, featuring four UAV-captured night scenes with calibrated poses.

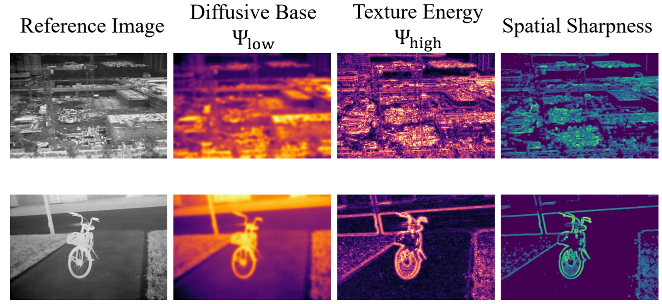

Figure 4: **Visualization of Decoupled Thermal Attributes. (a)** Reference Image. **(b)** Ψ_{low} models the smooth diffusive background. **(c)** Ψ_{high} captures high-frequency fine details. **(d)** The shape parameter map (Ψ_β) precisely aligns with geometric discontinuities, validating our physical truncation strategy.

Implementation Details. Our framework is implemented in PyTorch based on official 3DGS CUDA kernels. We train for $30,000$ iterations on a single NVIDIA RTX 3090 GPU. The spatial sharpness β is initialized to 2 and optimized with a learning rate of 0.001. We set the Gabor frequency bands $M = 4$. Loss weights are empirically set to $\lambda_c = 0.01$, $\lambda_s = 0.005$, and $\lambda_g = 0.01$. A $3,000$-iteration geometric warmup with standard 3DGS is applied before enabling spectral modulation. Input images are normalized to $[0, 1]$.

Baselines. We benchmark against three categories: (1) Standard 3DGS [Kerbl *et al.*, 2023]; (2) Thermal-Specific Methods, including Thermal3D-GS [Chen *et al.*, 2024], ThermalGaussian [Ye *et al.*, 2024], and NTR-Gaussian [Yang *et al.*, 2025]; and (3) Shape & Frequency-Aware Methods, comprising GES [Hamdi *et al.*, 2024], 3D-HGS [Li *et al.*, 2025], and WIPES [Zhang *et al.*, 2025a].

Metrics. Following the evaluation protocols established in previous works [Kerbl *et al.*, 2023; Chen *et al.*, 2024; Yang *et al.*, 2025], we report PSNR, SSIM, and LPIPS to quantitatively evaluate the reconstruction quality.

4.2 Comparisons

As detailed in Table 1, **S³**-TIR achieves consistent state-of-the-art performance across the RGBT-Scenes, TI-NSD, and NTR benchmarks, underscoring the robustness of our framework in diverse thermal environments. Most notably, on the RGBT-Scenes dataset—characterized by complex geometric and thermal distributions—our method delivers a substantial 42.5% **reduction in LPIPS** ($0.186 \rightarrow 0.107$) and a 3.2% improvement in PSNR compared to the second-best baseline, ThermalGaussian. This quantitative superiority is visually corroborated in Fig. 5, particularly in scenarios with significant emissivity variations, such as heated pipes and pedestrian silhouettes. While standard 3DGS and its thermal variants often yield over-smoothed artifacts, recent specialized methods exhibit distinct trade-offs: shape-aware ap-

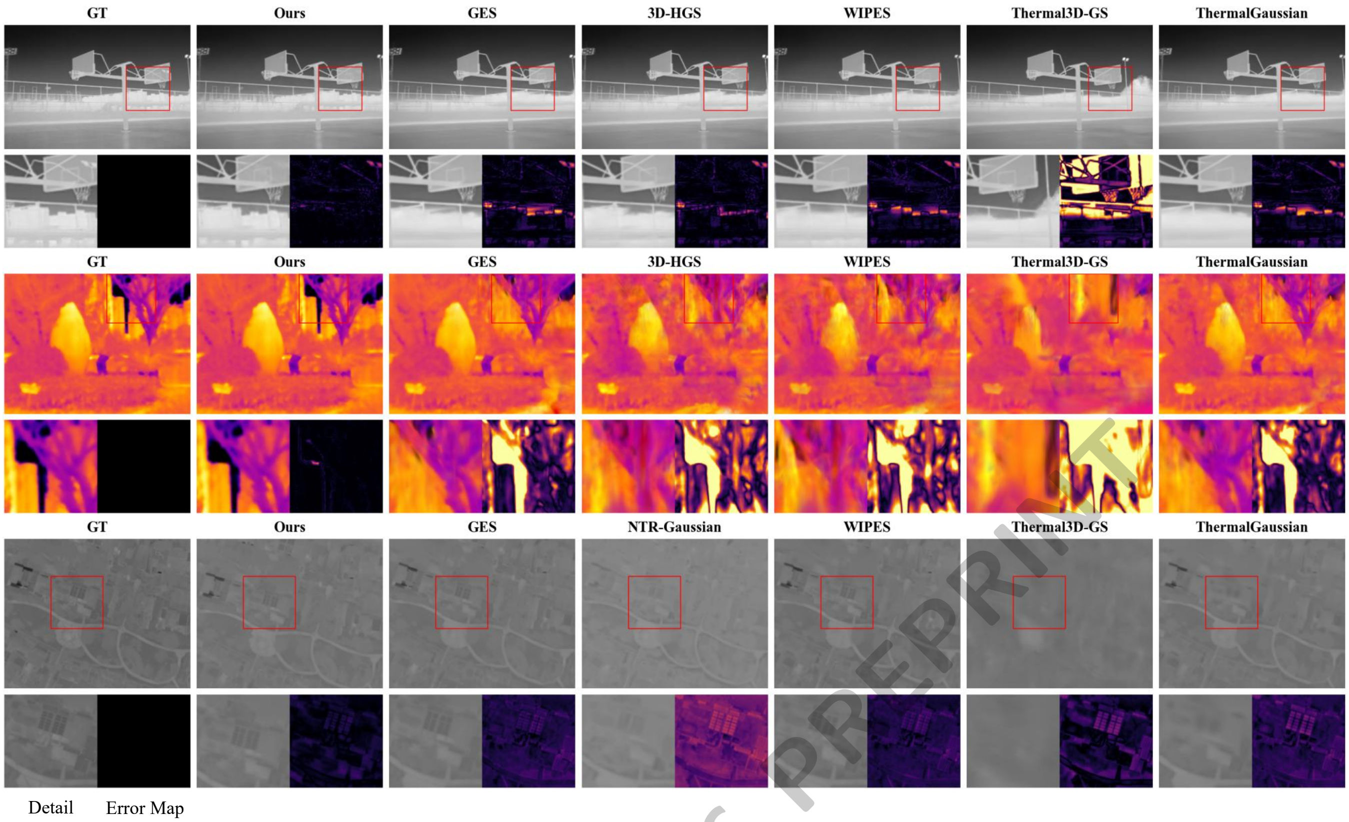

Figure 5: **Qualitative Comparisons on RGBT-Scenes, TI-NSD, and NTR Datasets.** We compare our method against state-of-the-art baselines, including GES , 3D-HGS , WIPES , Thermal3DGS , and NTR-Gaussian. The visualization of local details and corresponding error maps demonstrates that our method renders sharper edges and exhibits significantly less noise, achieving the highest fidelity relative to the ground truth (GT).

proaches like GES delineate sharp edges but lose internal gradient details, whereas frequency-aware methods like WIPES capture textures but suffer from boundary signal leakage. In contrast, $\mathbf{S^3}$-TIR achieves an optimal balance by leveraging generalized-exponential spatial support for precise physical truncation and enforcing multi-view frequency consistency. Consequently, our approach effectively mitigates non-physical high-frequency noise in background regions while accurately reconstructing the structural duality of diffusive smoothness and sharp structural discontinuities.

4.3 Ablation and Analysis

Contribution of Individual Components. We quantitatively validate the contribution of each component in Table 2. Specifically, removing the shape control mechanism (i.e., fixing $\beta \equiv 2$) results in a marked decline in SSIM, confirming that the generalized-exponential formulation is indispensable for modeling sharp thermal discontinuities. Similarly, disabling spectral modulation ($f = 0$) leads to over-smoothed reconstructions, evidenced by elevated LPIPS scores. This indicates that the model fails to capture fine-grained thermal textures arising from complex heat distributions without this component. In terms of optimization strategies, the most substantial performance drop occurs upon the removal of the multi-view frequency consistency loss ($\mathcal{L}_{consist}$). Absent this constraint, the optimization becomes susceptible to shape-texture ambiguity, resulting in view-dependent high-frequency jittering artifacts. Finally, the thermodynamic sparsity prior ($\mathcal{L}_{sparse}$) proves essential for noise suppression, ensuring that high-frequency modulation is selectively applied only where supported by physical evidence.

Effect of Spectral Modulation Bandwidth (M). The parameter M controls the spectral capacity for encoding internal thermal textures. As shown in Table 3, low bandwidths (e.g., $M = 1, 2$) constrain the representation to low-frequency diffusion, failing to capture fine-grained emissivity variations and resulting in over-smoothed artifacts. Conversely, excessive bands (e.g., $M = 8$) yield diminishing returns. Due to the inherent low Signal-to-Noise Ratio (SNR) of thermal sensors, over-parameterization tends to overfit sensor noise rather than physical structures. Empirically, $M = 4$ strikes the optimal balance, preserving sharp material textures while implicitly suppressing noise via band-limitation.

Analysis of Decoupled Attributes. To validate the physical fidelity of $\mathbf{S^3}$-TIR, we visualize the learned intermediate G-Buffers in Fig. 4. As shown in Fig. 4(b), the Low-Frequency Map Ψ_{low} captures a smooth, continuous distribution, characterizing macroscopic heat diffusion. In contrast,

Method	RGBT-Scenes			TI-NSD			NTR		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
3DGS	24.467	0.866	0.202	32.436	0.936	0.202	32.178	0.964	0.205
Thermal3D-GS	24.872	0.873	0.189	34.938	0.956	0.188	30.656	0.933	0.391
ThermalGaussian	25.086	0.876	0.186	32.429	0.891	0.224	32.061	0.963	0.234
NTR-Gaussian	-	-	-	-	-	-	27.765	0.939	0.263
GES	19.895	0.781	0.302	26.487	0.900	0.239	32.861	0.966	0.185
3D-HGS	20.770	0.794	0.298	34.140	0.955	0.201	29.491	0.951	0.353
WIPES	18.762	0.771	0.333	32.201	0.938	0.206	32.257	0.968	0.187
S³-TIR (Ours)	**25.877**	**0.945**	**0.107**	**35.071**	**0.955**	**0.194**	**33.181**	**0.973**	**0.181**

Table 1: **Quantitative Comparison on RGBT-Scenes, TI-NSD, and NTR.** Comparison of different methods for thermal infrared novel view synthesis. We compare our method against standard 3DGS, thermal-specific variants, and recent shape/frequency-aware primitives. For each dataset, we color the cells as best, second best, and third best.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Primitive Components			
w/o Shape Control ($\beta = 2$)	24.925	0.885	0.152
w/o Spectral Modulation ($\mathbf{f} = 0$)	24.613	0.892	0.168
Optimization Strategy			
w/o Consistency Loss ($\mathcal{L}_{consist} = 0$)	24.108	0.864	0.195
w/o Sparsity Prior ($\mathcal{L}_{sparse} = 0$)	25.534	0.938	0.115
S³-TIR (Ours)	**25.877**	**0.945**	**0.107**

Table 2: **Ablation Study.** Quantitative comparison of different primitive components and optimization strategies on the RGBT-Scenes dataset. We evaluate the contribution of Shape Control (β), Spectral Modulation ($\mathbf{f}$), Multi-View Consistency Loss ($\mathcal{L}_{consist}$), and Sparsity Prior ($\mathcal{L}_{sparse}$).

Bands (M)	PSNR ↑	SSIM ↑	LPIPS ↓
$M = 1$	33.420	0.915	0.238
$M = 2$	34.350	0.941	0.210
M = 4 (Ours)	**35.071**	**0.955**	**0.194**
$M = 6$	34.905	0.952	0.198
$M = 8$	34.680	0.948	0.203

Table 3: **Ablation on Gabor Frequency Bands** M (Evaluated on TI-NSD). We analyze the impact of spectral capacity on reconstruction quality on the TI-NSD dataset. $M = 4$ consistently provides the optimal balance between texture richness and noise suppression.

the High-Frequency Map Ψ_{high} (Fig. 4c) exhibits structural sparsity, selectively encoding details arising from material heterogeneity (e.g., windows). Crucially, the Spatial Sharpness Map Ψ_{β} (Fig. 4d) strictly delineates geometric silhouettes. High β values align with object boundaries, confirming that our Generalized-Exponential support enforces physical truncation to prevent signal leakage. This clear separation substantiates that our representation successfully resolves the ambiguity between geometry and texture.

Validation of Structural Duality To validate our Coupled Spatial-Spectral representation, we analyze the distribution of the shape parameter β. As shown in Fig. 6, thermal scenes exhibit a distinct *structural duality*. Spatially (Fig. 6a), high β values concentrate at material discontinuities to enforce boundary truncation, whereas low β values dominate the diffusive background. Quantitatively, Fig. 6b confirms a *structured sparsity*: a heavy-tailed distribution where most primitives follow Gaussian diffusion ($\beta \approx 2$), while a sparse subset requires generalized-exponential truncation ($\beta \gg 2$). This validates our learnable shape parameter. In contrast, standard fixed kernels fail to capture this duality, causing the over-smoothed edges observed in baselines.

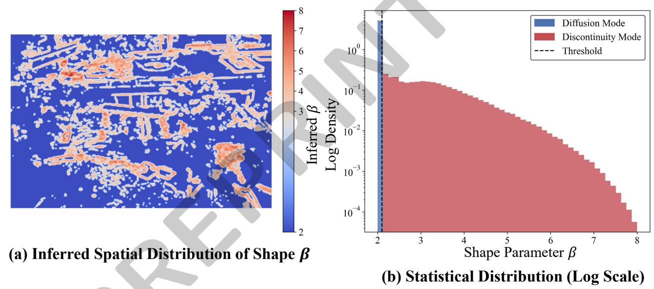

Figure 6: **Analysis of Thermal Shape Parameter** β**. (a) Spatially**, high β values (red) strictly localize at structural boundaries (*Discontinuity Mode*), separating from the smooth diffusive background (blue). **(b) Statistically**, the log-histogram confirms **structured sparsity**: a heavy-tailed distribution dominated by Gaussian diffusion ($\beta \approx 2$) with sparse edge truncation ($\beta \gg 2$).

5 Conclusion

We present **S³-TIR**, a physics-inspired framework that resolves the inherent duality of thermal infrared novel view synthesis. By coupling adaptive spatial support with internal spectral modulation via a unified generalized-exponential Gabor primitive, our method confines high-frequency textures within geometric boundaries while preserving diffusive smoothness. Our frequency-aware deferred splatting pipeline further resolves decomposition ambiguity by enforcing multi-view consistency on decoupled frequency attributes. Extensive experiments demonstrate that **S³**-TIR achieves state-of-the-art performance, generating structurally stable thermal sequences free from high-frequency jittering.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62372382).

References

[Bao *et al.*, 2023] 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.

[Bao *et al.*, 2025] Liuxin Bao, Xiaofei Zhou, Bolun Zheng, Runmin Cong, et al. Ifenet: Interaction, fusion, and enhancement network for vdt salient object detection. *TIP*, 2025.

[Chen *et al.*, 2004] Thou-Ho Chen, Ping-Hsueh Wu, and Yung-Chuen Chiou. An early fire-detection method based on image processing. In *ICIP*, volume 3, pages 1707–1710. IEEE, 2004.

[Chen *et al.*, 2024] Qian Chen, Shihao Shu, and Xiangzhi Bai. Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis. In *ECCV*, pages 253–269. Springer, 2024.

[Chen *et al.*, 2025] Jincheng Chen, Dengfeng Ren, Yuge Han, and Qunqing Lin. A novel framework for predicting 3d scene infrared radiation characteristics through ai-enhanced thermodynamic modeling. *Int. J. Heat Mass Transf.*, 238:126434, 2025.

[Di *et al.*, 2024] Yan Di, Chenyangguang Zhang, Chaowei Wang, Ruida Zhang, Guangyao Zhai, Yanyan Li, Bowen Fu, Xiangyang Ji, and Shan Gao. Shapematcher: Self-supervised joint shape canonicalization segmentation retrieval and deformation. In *CVPR*, pages 21017–21028, 2024.

[Gao *et al.*, 2024] Jian Gao, Chun Gu, Youtian Lin, Zhihao Li, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing. In *ECCV*, pages 73–89. Springer, 2024.

[Hahn and Özisik, 2012] David W Hahn and M Necati Özisik. *Heat conduction*. John Wiley & Sons, 2012.

[Hamdi *et al.*, 2024] Abdullah Hamdi, Luke Melas-Kyriazi, Jinjie Mai, Guocheng Qian, Ruoshi Liu, Carl Vondrick, Bernard Ghanem, and Andrea Vedaldi. Ges: Generalized exponential splatting for efficient radiance field rendering. In *CVPR*, pages 19812–19822, 2024.

[Hassan *et al.*, 2024] Mariam Hassan, Florent Forest, Olga Fink, and Malcolm Mielle. Thermonerf: Joint rgb and thermal novel view synthesis for building facades using multimodal neural radiance fields. *arXiv*, 2024.

[Hu *et al.*, 2025] Yufan Hu, Zekai Shao, Bin Fan, and Hongmin Liu. Dual-level modality de-biasing for rgb-t tracking. *TIP*, 2025.

[Huang *et al.*, 2024] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2D Gaussian splatting for geometrically accurate radiance fields. In *SIGGRAPH*, pages 1–11, 2024.

[Huang *et al.*, 2025] Yi-Hua Huang, Ming-Xian Lin, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Deformable radial kernel splatting. In *CVPR*, pages 21513–21523, 2025.

[Jiang *et al.*, 2024] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *CVPR*, pages 5322–5332, 2024.

[Kerbl *et al.*, 2023] 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.

[Kheradmand *et al.*, 2024] Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Weiwei Sun, Yang-Che Tseng, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, and Kwang Moo Yi. 3d gaussian splatting as markov chain monte carlo. *NeurIPS*, 37:80965–80986, 2024.

[Li *et al.*, 2025] Haolin Li, Jinyang Liu, Mario Sznaier, and Octavia Camps. 3d-hgs: 3d half-gaussian splatting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 10996–11005, 2025.

[Liang *et al.*, 2024] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. In *CVPR*, pages 21644–21653, 2024.

[Lin *et al.*, 2026] Ling Lin, Yaoxing Wang, Congcong Zhu, and Jingrun Chen. A2net: Affiliation alignment networks for whole-body pose estimation with vision–language models. *TNNLS*, 2026.

[Lin *et al.*, 2all] Ling Lin, Hao Liu, Jinqiao Liang, Zhendong Li, Jiao Feng, and Hu Han. Consensus-agent deep reinforcement learning for face aging. *IEEE Trans. Image Process.*, 33:1795–1809, 2all.

[Liu *et al.*, 2024] Yuxiang Liu, Rouwan Wu, Shen Yan, Xiaoya Cheng, Juelin Zhu, Yu Liu, and Maojun Zhang. Atloc: Aerial thermal images localization via view synthesis. *IEEE Trans. Geosci. Remote Sens.*, 62:1–13, 2024.

[Liu *et al.*, 2025a] Rong Liu, Dylan Sun, Meida Chen, Yue Wang, and Andrew Feng. Deformable beta splatting. In *SIGGRAPH*, pages 1–11, 2025.

[Liu *et al.*, 2025b] Yuxiang Liu, Xi Chen, Shen Yan, Zeyu Cui, Huaxin Xiao, Yu Liu, and Maojun Zhang. Thermalgs: Dynamic 3d thermal reconstruction with gaussian splatting. *Remote Sensing*, 17(2):335, 2025.

[Lu *et al.*,] Rongfeng Lu, Hangyu Chen, Zunjie Zhu, Yuhang Qin, Ming Lu, Chenggang Yan, et al. Thermalgaussian: Thermal 3d gaussian splatting. In *ICLR*.

[Ma *et al.*, 2023] Mei Ma, Ling Lin, Heng Wang, Zhendong Li, and Hao Liu. Cross-modality fourier feature for medical image synthesis. In *ICME*, pages 1475–1480. IEEE, 2023.

[Narasimhan, 1999] Thiruppudaimarudhur N Narasimhan. Fourier's heat conduction equation: History, influence, and connections. *Rev. Geophys.*, 37(1):151–172, 1999.

[Rangel *et al.*, 2014] Johannes Rangel, Samuel Soldan, and A Kroll. 3d thermal imaging: Fusion of thermography and depth cameras. In *QIRT*, volume 3, 2014.

[Ren *et al.*, 2025] Yufan Ren, Zicong Jiang, Tong Zhang, Søren Forchhammer, and Sabine Süsstrunk. Fds: Frequency-aware denoising score for text-guided latent diffusion image editing. In *CVPR*, pages 2651–2660, 2025.

[Ring and Ammer, 2012] EFJ Ring and Kurt Ammer. Infrared thermal imaging in medicine. *Physiol. Meas.*, 33(3):R33, 2012.

[Shi *et al.*, 2023] Jiangming Shi, Yachao Zhang, Xiangbo Yin, Yuan Xie, Zhizhong Zhang, Jianping Fan, Zhongchao Shi, and Yanyun Qu. Dual pseudo-labels interactive self-training for semi-supervised visible-infrared person re-identification. In *ICCV*, pages 11218–11228, 2023.

[Shi *et al.*, 2024] Jiangming Shi, Xiangbo Yin, Yeyun Chen, Yachao Zhang, Zhizhong Zhang, Yuan Xie, and Yanyun Qu. Multi-memory matching for unsupervised visible-infrared person re-identification. In *ECCV*, pages 456–474. Springer, 2024.

[Song *et al.*, 2024] Gaochao Song, Chong Cheng, and Hao Wang. Gvkf: Gaussian voxel kernel functions for highly efficient surface reconstruction in open scenes. *NeurIPS*, 37:104792–104815, 2024.

[Vidas and Moghadam, 2013] Stephen Vidas and Peyman Moghadam. Heatwave: A handheld 3d thermography system for energy auditing. *Energy Build.*, 66:445–460, 2013.

[Wang *et al.*, 2023] Yadong Wang, Darui Jin, Junzhang Chen, and Xiangzhi Bai. Revelation of hidden 2d atmospheric turbulence strength fields from turbulence effects in infrared imaging. *Nat. Comput. Sci.*, 3(8):687–699, 2023.

[Wang *et al.*, 2025a] Dongqing Wang, Ehsan Pajouheshgar, Yitao Xu, Tong Zhang, and Sabine Süsstrunk. Volumetric temporal texture synthesis for smoke stylization using neural cellular automata. *arXiv preprint arXiv:2502.09631*, 2025.

[Wang *et al.*, 2025b] Qiang Wang, Shiyu Guan, Shuchang Lyu, and Guangliang Cheng. Enhancing drone-based fire detection with flame-specific attention and optimized feature fusion. *Int. J. Appl. Earth Obs. Geoinf.*, 142:104655, 2025.

[Wang *et al.*, 2025c] Zhaozhi Wang, Yue Liu, Yunjie Tian, Yunfan Liu, Yaowei Wang, and Qixiang Ye. Building vision models upon heat conduction. In *CVPR*, pages 9707–9717, 2025.

[Wang *et al.*, 2026] Zhongwen Wang, Han Ling, Weihao Zhang, Yinghui Sun, and Quansen Sun. Etgs: Explicit thermodynamics gaussian splatting for dynamic thermal reconstruction. In *ICLR*, 2026.

[Widder, 1976] David Vernon Widder. *The heat equation*, volume 67. Academic Press, 1976.

[Wurster *et al.*, 2024] Skylar Wurster, Ran Zhang, and Changxi Zheng. Gabor splatting for high-quality gigapixel image representations. In *SIGGRAPH Posters*, pages 1–2. 2024.

[Yang *et al.*, 2025] Kun Yang, Yuxiang Liu, Zeyu Cui, Yu Liu, Maojun Zhang, Shen Yan, and Qing Wang. Ntr-gaussian: Nighttime dynamic thermal reconstruction with 4d gaussian splatting based on thermodynamics. In *CVPR*, pages 691–700, 2025.

[Ye *et al.*, 2024] 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 *IROS*, pages 1046–1053. IEEE, 2024.

[Yu *et al.*, 2024] Zehao Yu, Torsten Sattler, and Andreas Geiger. Gaussian opacity fields: Efficient adaptive surface reconstruction in unbounded scenes. *ACM Trans. Graph.*, 43(6):1–13, 2024.

[Zeng *et al.*, 2025] Zhaojie Zeng, Yuesong Wang, Lili Ju, and Tao Guan. Frequency-aware density control via reparameterization for high-quality rendering of 3d gaussian splatting. In *AAAI*, volume 39, pages 9833–9841, 2025.

[Zhang *et al.*, 2024a] Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, and Eric Xing. Fregs: 3d gaussian splatting with progressive frequency regularization. In *CVPR*, pages 21424–21433, 2024.

[Zhang *et al.*, 2024b] Ruida Zhang, Chenyangguang Zhang, Yan Di, Fabian Manhardt, Xingyu Liu, Federico Tombari, and Xiangyang Ji. Kp-red: Exploiting semantic keypoints for joint 3d shape retrieval and deformation. In *CVPR*, pages 20540–20550, 2024.

[Zhang *et al.*, 2025a] Wenhao Zhang, Hao Zhu, Delong Wu, Di Kang, Linchao Bao, Xun Cao, and Zhan Ma. Wipes: Wavelet-based visual primitives. In *ICCV*, pages 27338–27347, 2025.

[Zhang *et al.*, 2025b] Xiaoyu Zhang, Weihong Pan, Chong Bao, Xiyu Zhang, Xiaojun Xiang, Hanqing Jiang, and Hujun Bao. Lookcloser: Frequency-aware radiance field for tiny-detail scene. In *CVPR*, pages 16122–16132, 2025.

[Zhao *et al.*, 2017] Shibo Zhao, Zheng Fang, and Shiguang Wen. A real-time handheld 3d temperature field reconstruction system. In *IEEE CYBER*, pages 289–294. IEEE, 2017.

[Zhao *et al.*, 2025a] Beizhen Zhao, Yifan Zhou, Sicheng Yu, Zijian Wang, and Hao Wang. Wavelet-gs: 3d gaussian splatting with wavelet decomposition. In *ACM MM*, pages 8616–8625, 2025.

[Zhao *et al.*, 2025b] Chen Zhao, Xuan Wang, Tong Zhang, Saqib Javed, and Mathieu Salzmann. Self-ensembling gaussian splatting for few-shot novel view synthesis. In *ICCV*, pages 4940–4950, 2025.

[Zhou *et al.*, 2025] Junyu Zhou, Yuyang Huang, Wenrui Dai, Junni Zou, Ziyang Zheng, Nuowen Kan, Chenglin Li, and Hongkai Xiong. 3dgabsplat: 3d gabor splatting for frequency-adaptive radiance field rendering. In *ACM MM*, pages 72–81, 2025.