SupIR-GS: Thermal Infrared Super-Resolution Novel View Synthesis with Imaging-Calibrated 3D Gaussian Splatting

Jin Liu[1,2*], Haodong Li[1,2*], Jiagang Chen[3], Dabin Leng[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. Thermal infrared (TIR) imaging facilitates perception, security surveillance, and industrial inspection. Compared with visible spectrum, TIR observations are typically resolution-limited and texture-scarce. More critically, thermal diffusion smooths true geometric boundaries into low-frequency blur, whereas non-uniform sensor response and fixed-pattern noise introduce structured high-frequency artifacts, yielding a pronounced signal-to-noise inversion. Consequently, reconstructing a high-resolution (HR) 3D TIR scene and synthesizing high-quality novel views from only low-resolution (LR) infrared inputs remains challenging. To address this, we propose the first imaging-calibrated TIR super-resolution 3D Gaussian Splatting (3DGS) framework that reconstructs high-resolution 3D TIR scenes using only LR inputs. Specifically, we explicitly model infrared imaging degradation on the clean radiation signals rendered by 3DGS using physics-informed priors, enabling decoupling of the underlying thermal radiation from sensor-induced artifacts. To alleviate texture scarcity, an intensity-conditioned tone adapter is designed to apply local affine residual modulation, thereby surfacing latent thermal gradients. Finally, we implement a frequency-aware curriculum learning strategy that leverages edge priors from a super-resolution teacher network, and employs a cosine ramp-up schedule to smoothly shift the optimization focus from coarse low-frequency structures to fine high-frequency details. Extensive evaluations across TIR benchmarks show that our method significantly outperforms existing works in both high-resolution synthesis quality and quantitative temperature fidelity.

Keywords: Thermal Infrared Imaging · Super-Resolution · Novel View Synthesis · 3D Gaussian Splatting

* Equal Contribution.
† Corresponding Author.

1 Introduction

Thermal infrared (TIR) imaging measures long-wave infrared radiation to characterize the thermal signatures of a scene, providing stable observations under darkness, smoke, and other visibility-degraded conditions. Such robustness supports all-weather perception across diverse applications, including autonomous driving, disaster response, and industrial inspection [7, 12, 23, 28]. However, constrained by sensor manufacturing costs and the physical diffraction limit, off-the-shelf thermal cameras typically yield low-resolution (LR) observations [13,17,25]. This lack of spatial resolution limits the use of TIR data in fine-grained 3D scene reconstruction and immersive rendering, making high-fidelity infrared synthesis from low-resolution observations a standing challenge in computer vision and remote sensing [26, 39].

Recent advances in 3D scene representations, exemplified by neural radiance fields (NeRF) [1, 21, 29] and 3D Gaussian Splatting (3DGS) [6, 14, 36], have achieved remarkable success in the visible spectrum. However, their direct transfer to thermal infrared imagery [3, 19, 31] is challenged by a non-trivial geometry-radiance ambiguity rooted in infrared-specific physical degradations. Specifically, thermal diffusion and atmospheric scattering tend to attenuate true physical boundaries into low-contrast, low-frequency cues. Conversely, uncooled microbolometer artifacts, such as Non-Uniformity Response (NUR) and fixed pattern noise (FPN), often manifest as structured high-frequency textures. This signal-to-noise ratio inversion biases standard 3DGS optimization toward suboptimal solutions, where high-frequency floater artifacts emerge to explain sensor noise, or geometry collapses under insufficient edge supervision. Such issues are exacerbated under low-resolution inputs, severely impeding the stable recovery of sharp boundaries and view-consistent thermal radiance fields.

Addressing these challenges, we present a super-resolution Gaussian Splatting framework (SupIR-GS) for novel-view synthesis from low-resolution multi-view TIR images. Rather than fitting 3DGS directly to raw observations, a latent high-resolution rendering is first produced by 3DGS and subsequently align into LR observations through differentiable physical calibration and intensity modulation. By enforcing supervision in this aligned domain, geometric optimization is primarily driven by cross-view consistent structural cues, while the influence of spurious high-frequency sensor components on gradient updates is reduced.

Specifically, we apply a Physically-constrained Infrared Calibration module (PhysIR-Cal) to the clean radiation signals rendered by 3DGS, parameterized as a differentiable layer to suppress high-frequency artifacts from non-uniformity and fixed pattern noise. To address narrow dynamic range of TIR imagery, an Intensity-conditioned Tone Adapter (IRTone-Adapter) applies local affine residual modulation, adaptively enhancing contrast in texture-scarce regions. Furthermore, we develop a Frequency-Aware Curriculum Learning strategy to stabilize optimization. By leveraging edge prior from a pretrained super-resolution teacher, it guides the reconstruction from a low-frequency geometric scaffold to high-frequency details. Extensive evaluations on real-world datasets, including TI-NSD and ThermoScenes, demonstrate that proposed SupIR-GS achieves superior high-resolution novel-view synthesis and improved geometric fidelity over strong baselines.

The main contributions are summarized as follows:

1. To the best of our knowledge, we are the first to propose SupIR-GS, an infrared super-resolution 3DGS framework inspired by imaging calibration. It enables 3D thermal scene reconstruction with high fidelity using only low-resolution inputs by bridging the gap between latent radiance fields and degraded observations.
2. We establish a differentiable forward-imaging scheme incorporating PhysIR-Cal for physics-informed degradation calibration and IRTone-Adapter for local intensity modulation, thereby aligning latent 3DGS outputs with the degraded observation space to resolve imaging discrepancies.
3. We develop a Frequency-Aware Curriculum Learning strategy that leverages structural priors to guide a stable optimization trajectory from coarse structures to fine-grained details, while suppressing spurious high frequency artifacts induced by sensor noise.

2 Related Work

2.1 Novel View Synthesis

NeRF [21] enable high-quality novel view synthesis by optimizing an implicit volumetric scene representation, while subsequent works improve anti-aliasing and multiscale rendering behavior (e.g., Mip-NeRF) [1]. To further accelerate training and rendering, Instant-NGP [22] introduces hash-grid encodings for fast convergence. Recently, 3D GS [14] achieves real-time radiance field rendering by representing scenes with explicit anisotropic Gaussian primitives and differentiable rasterization. Follow-up Gaussian variants improve aliasing robustness [36], view-adaptive geometry [20], and surface-aligned reconstruction [8], while compression-oriented methods [5] further reduce storage and boost throughput. Despite strong performance on RGB imagery, directly transferring these pipelines to TIR imagery remains challenging. Unlike visible-band signals, TIR observations are strongly affected by sensor-specific degradations, especially FPN and NUR. These structured artifacts introduce geometric-radiometric ambiguity during optimization, where high-frequency noise patterns can be mistakenly interpreted as physical edges and fine geometry.

2.2 Thermal Infrared Novel View Synthesis

Early thermal novel view synthesis often relied on multi-sensor fusion [26] or online temperature field reconstruction systems [39]. However, thermal imagery typically exhibits low texture and weak repeatable gradients, making feature matching brittle and limiting classical SfM-style pipelines. More recently, implicit representations have been adapted to the thermal domain. Thermal-NeRF [31] reconstructs radiance fields by introducing structural constraints suited for TIR observations, while multimodal formulations further leverage cross-spectral cues for improved stability and fidelity, such as ThermoNeRF [9] and ThermalNeRF [18]. To enable real-time thermal NVS, explicit Gaussian-based frameworks have also been explored, including ThermalGaussian [19] and physics-induced Thermal3D-GS [3]. Nevertheless, many existing thermal pipelines implicitly assume well-calibrated inputs and continue to rely on standard photometric supervision. As a result, FPN- and NUR-related artifacts may be absorbed into geometry or appearance parameters during optimization, producing spurious high-frequency structure and reconstruction collapse in challenging scenes.

2.3 Thermal Infrared Super-Resolution

Thermal imaging, especially with uncooled microbolometers, is inherently constrained by hardware limits, resulting in low spatial resolution and prominent structured noise. A line of work studies thermal/infrared single-image super-resolution (SR), including challenge settings and dedicated architectures [2,4,10, 13,17,25]. While these 2D SR methods can enhance perceptual sharpness, they do not enforce multi-view consistency and often hallucinate view-inconsistent pseudo-textures, which can severely destabilize downstream 3D optimization.

Beyond 2D thermal SR, recent works have explored combining image restoration priors with 3D rendering optimization. For NeRF-based rendering, Cross-Guided Optimization [33] integrates multi-view image super-resolution with radiance-field training, and GMT [32] improves generalizable neural rendering via geometry-driven multi-reference texture transfer. Recent 3DGS-based restoration methods further address degraded observations, including deblurring [15], degradation-agnostic feed-forward reconstruction [34], and arbitrary-scale Gaussian super-resolution [37], alongside Gaussian-based 3D SR methods [6,27,35].

However, most of these methods are designed for RGB imagery. Their objectives often assume that high-frequency gradients correspond to true geometric or textural details. This assumption is ill-suited for TIR data, where thermal diffusion smooths physical boundaries into low-frequency transitions, while FPN and NUR introduce structured high-frequency artifacts. Therefore, directly applying RGB restoration to thermal imagery may interpret hardware-induced patterns as geometry.

3 The Proposed SupIR-GS

Given multi-view LR thermal infrared sequences $\{I_v^{\mathrm{lr}}\}$, the proposed SupIR-GS learns a 3DGS representation optimized for super-resolution novel-view synthesis. As illustrated in Fig. 1, we natively render an ideal TIR image $\hat{I}_v$ at the target HR setting . This rendering is mapped to the observation domain via a differentiable imaging chain: PhysIR-Cal parameterizes systematic degradations (e.g., NUR and FPN), and IRTone-Adapter applies intensity-conditioned local intensity modulation. By aligning these processed predictions with LR observations, the supervision is backpropagated to optimize the fundamental Gaussian

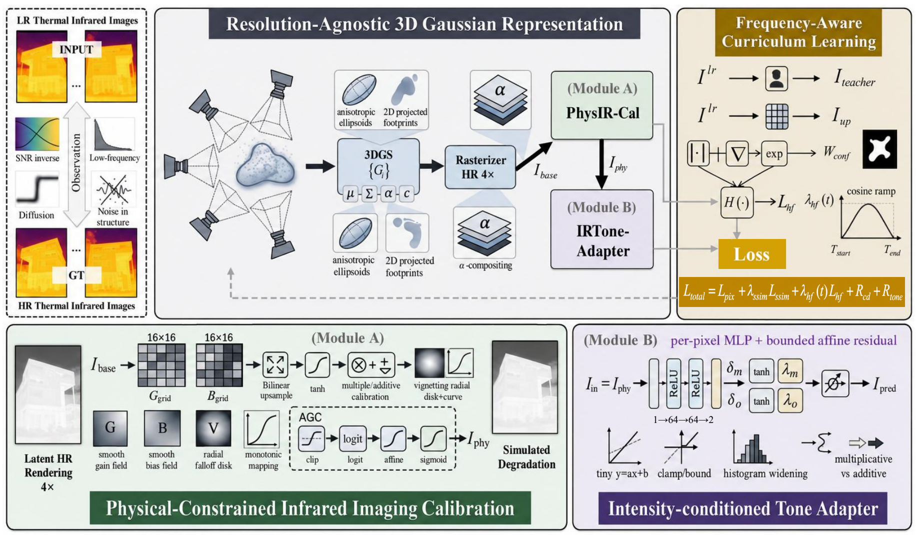

Fig. 1: Overview of the proposed super-resolution Gaussian reconstruction framework. The pipeline incorporates a differentiable PhysIR-Cal layer for degradation modeling, an IRTone-Adapter for contrast modulation, and a Frequency-Aware Curriculum Learning strategy to stabilize high-frequency detail recovery.

attributes (including positions, scales, covariances, and spherical harmonic coefficients), intrinsically embedding high-frequency details into 3D geometry rather than relying on image-space post-processing. Finally, a Frequency-Aware Curriculum Learning strategy guides the optimization from structural scaffolds to fine-grained details, preserving geometric consistency against the challenges of intrinsic thermal diffusion.

3.1 Resolution-Agnostic 3D Gaussian Representation

3DGS represents a scene as a set of anisotropic 3D Gaussian primitives $\{G_i\}_{i=1}^{N}$. Each primitive is parameterized by a 3D mean $\mu_i \in \mathbb{R}^3$ and a covariance $\Sigma_i \in \mathbb{R}^{3\times 3}$, with spatial density

$$G_i(\mathbf{x}) = \exp\Big(-\frac{1}{2}(\mathbf{x}-\mu_i)^\top \Sigma_i^{-1}(\mathbf{x}-\mu_i)\Big). \tag{1}$$

To enforce the symmetry and positive definiteness of Σ_i while facilitating optimization, 3DGS adopts a rotation–scale parameterization,

$$\Sigma_i = R_i\, S_i^2\, R_i^\top, \qquad S_i = \mathrm{diag}(s_i),\ s_i \in \mathbb{R}_+^3, \tag{2}$$

where $R_i \in SO(3)$ denotes a rotation matrix and s_i denotes the per-axis scales.

For a given viewpoint v, these 3D Gaussians are projected onto the 2D image plane. The synthesized intensity at pixel $\mathbf{p}$ is computed via a tile-based differentiable rasterizer using α-compositing:

$$\hat{I}_v(\mathbf{p}) = \sum_{i\in\mathcal{N}(\mathbf{p})} c_i \alpha_i g_i(\mathbf{p}) \prod_{j<i} (1 - \alpha_j g_j(\mathbf{p})), \tag{3}$$

where $g_i(\mathbf{p})$ denotes the screen-space footprint of the i-th Gaussian, c_i is its associated radiance (a single-channel intensity for TIR), and $\mathcal{N}(\mathbf{p})$ is the set of sorted Gaussians overlapping pixel $\mathbf{p}$.

Since the Gaussian primitives $\{G_i\}$ are defined as continuous functions in 3D space, the underlying radiance field is resolution-agnostic by construction. To synthesize the super-resolved infrared image $\hat{I}_v^{hr}$, we calculate Eq. (3) on a high-density sampling grid Ω_{hr} with r^2 times more query points than the LR observation (where r is the upsampling factor). This enables the model to natively recover fine-grained thermal signatures by querying the continuous primitives at an arbitrary spatial frequency.

3.2 Physical-Constrained Infrared Imaging Calibration

Uncooled microbolometer observations exhibit a pronounced signal-to-noise ratio inversion: FPN and NUR effects typically appear as sharp, structured high-frequency textures, whereas the underlying thermal radiance field is blurred by thermal diffusion. This spectral misalignment biases standard 3DGS toward fitting hardware noise as spurious geometry, leading to unstable primitive densification and persistent floater artifacts.

Addressing physics-induced geometric degradation, PhysIR-Cal functions as a differentiable forward calibration layer between the radiance-domain rendering and raw observations. This module explicitly models systematic imaging degradations to decouple scene radiance from device-dependent responses, thereby shielding Gaussian optimization from spurious high-frequency gradients.

We eschew direct pixel-wise noise optimization for a parameter-efficient calibration field to prevent overfitting under low SNR. Specifically, for an ideal radiance image I_{base} rendered by 3DGS, we define two learnable low-resolution grids $G_{\text{grid}}, B_{\text{grid}} \in \mathbb{R}^{16\times16}$, which are bilinearly upsampled to the target resolution. Bounded by a $\tanh(\cdot)$ nonlinearity, these grids generate spatial gain (G) and bias (B) fields to parameterize the multiplicative and additive sensor deviations inherent in NUR and FPN.

A radial vignetting model is employed to address lens-induced photometric attenuation, defined as a function of the normalized radial distance r:

$$V(r) = \exp\left(a_2 r^2 + a_4 r^4\right), \tag{4}$$

in which a_2 and a_4 are learnable coefficients that parameterize the falloff profile.

In addition, since thermal cameras may apply per-frame histogram stretching or automatic gain control (AGC), we optionally employ a view-dependent monotonic tone mapping formulated in the logit domain to absorb global intensity re-scaling across frames. Specifically, when the AGC module is enabled, we apply

$$I_{\text{phy}} = \sigma\left(s_k \cdot \operatorname{logit}\left(\tilde{I}_{\text{calib}}\right) + t_k\right), \tag{5}$$

where $\tilde{I}_{\text{calib}} = \operatorname{clip}(I_{\text{calib}}, \varepsilon, 1-\varepsilon)$ is a numerically stabilized version of I_{calib} with a small $\varepsilon > 0$. The calibrated radiance before the optional tone mapping is

$$I_{\text{calib}} = \left[I_{\text{base}} \cdot (1 + G) + B\right] \cdot V, \tag{6}$$

where s_k, t_k denotes the gain and bias for view k, $\sigma(\cdot)$ is the sigmoid function, and $\operatorname{logit}(x) = \ln\left(\frac{x}{1-x}\right)$ is the inverse of $\sigma(\cdot)$. When the AGC module is disabled, we simply set $I_{\text{phy}} = \tilde{I}_{\text{calib}}$.

Under this physically constrained decoupling, PhysIR-Cal acts as a learnable low-frequency calibration layer during optimization, attenuating high-frequency distractions induced by FPN/NUR. As a result, densification and splitting of 3D Gaussian primitives are biased toward signals consistent with scene structure across views, providing a more reliable training signal for subsequent stable geometry recovery and novel-view synthesis.

3.3 Intensity-conditioned Tone Adapter

Unlike visible imagery, thermal infrared observations are often constrained by a narrow dynamic range and flattened intensity distributions. Consequently, the high-frequency variations critical for multi-view consistency become significantly attenuated. In such texture-scarce regions, pixel-wise reprojection errors yield unstable gradients for refining fine-scale geometry. To address this limitation, we introduce the IRTone-Adapter, a lightweight, intensity-conditioned residual module. Acting as a locally adaptive tone mapper, it is designed to surface latent thermal variations from otherwise flat infrared measurements.

Specifically, IRTone-Adapter is implemented as a per-pixel MLP Φ_{tone} with two 64-channel hidden layers. For each spatial location p, the network takes the rendered intensity $I_{\text{in}}(p)$ to predict constrained local affine residuals: a multiplicative gain δ_m and an additive bias δ_o. To ensure training stability while limiting appearance capacity, these outputs are range-bounded via $\tanh(\cdot)$ and modulated by preset scaling factors:

$$[\delta_m, \delta_o] = \Phi_{\text{tone}}(\operatorname{mean}(I_{\text{in}}(p)); \theta_{\text{tone}}). \tag{7}$$

$$I_{\text{pred}}(p) = I_{\text{in}}(p) \cdot \left(1 + \lambda_m \tanh(\delta_m)\right) + \lambda_o \tanh(\delta_o). \tag{8}$$

Here, δ_m and δ_o denote the predicted residuals for multiplicative gain and additive bias, respectively. In our implementation, we set $\lambda_m = 0.08$ and $\lambda_o = 0.04$, which bound the multiplicative and additive residuals to ± 0.08 and ± 0.04, respectively. Such a constrained formulation restricts the adapter to small corrections around the underlying thermal radiance, reducing the possibility of masking geometric errors through large intensity remapping.

3.4 Frequency-Aware Curriculum Learning Strategy

Thermal diffusion and atmospheric scattering inherently blur physical boundaries, yielding TIR imagery with weak edge contrast. Under low-resolution observations, high-frequency cues become ambiguous: some correspond to genuine thermal boundaries, while others originate from FPN/NUR artifacts or upsampling noise. Therefore, our objective is not to blindly enhance all high-frequency responses, but to retain reliable high-frequency cues that are consistent with the low-frequency observation and the multi-view optimization process. To address this trade-off, we formulate a frequency-aware curriculum learning strategy. By leveraging structural priors from a pre-trained SwinIR teacher, we exploit the spectral bias in training dynamics to guide the optimization from low-frequency structure toward high-frequency detail over iterations.

A confidence mask W_{conf} is first defined to assess the reliability of teacher-provided details from pixel discrepancies and gradient consistency. Reliability at location $\mathbf{p}$ is assessed by measuring the joint deviation between the teacher prediction I_{teacher} and the bilinearly upsampled observation I_{up} in both intensity and gradient domains:

$$W_{\text{conf}}(\mathbf{p}) = \exp\Big(-\gamma_p\big|I_{\text{teacher}}(\mathbf{p}) - I_{\text{up}}(\mathbf{p})\big| - \gamma_g\big\|\nabla I_{\text{teacher}}(\mathbf{p}) - \nabla I_{\text{up}}(\mathbf{p})\big\|_1\Big), \quad (9)$$

where $W_{\text{conf}}(\mathbf{p})$ is a scalar weight, ∇ denotes a spatial gradient operator, and $\|\cdot\|_1$ is the ℓ_1 norm; γ_p and γ_g control the sensitivity of confidence decay to pixel and gradient errors. The exponential mapping yields weights in $(0, 1]$, approaching unity only when the teacher output matches the observation in low-frequency structure, under which the associated high-frequency details are treated as more reliable.

Based on this mask, a teacher-guided high-frequency residual constraint is imposed. Let $\hat{I}_v$ be the current rendering. Using a high-pass filter $H(\cdot)$ (implemented as differences of average pooling), we define

$$\mathcal{L}_{\text{hf}} = \frac{1}{N}\sum_{\mathbf{p}} W_{\text{conf}}(\mathbf{p})\,\big|H(I_{\text{pred}}(\mathbf{p})) - H(I_{\text{teacher}}(\mathbf{p}))\big|. \quad (10)$$

The weight $W_{\text{conf}}(\mathbf{p})$ enforces high-frequency matching only in teacher-reliable regions, mitigating spurious details in textureless or inconsistent areas.

To enable stage-wise optimization, we progressively activate the high-frequency term with a cosine ramp. Early iterations focus on photometric consistency and SSIM losses to form a stable low-frequency scaffold; later, the weight $\lambda_{hf}(t)$ increases to emphasize edge details in confident regions:

$$\lambda_{\text{hf}}(t) = \begin{cases} 0, & t \leq T_{\text{start}} \\ \frac{\lambda_{\text{hf}}^{\max}}{2}\Big[1 - \cos\Big(\pi\frac{t - T_{\text{start}}}{T_{\text{end}} - T_{\text{start}}}\Big)\Big], & T_{\text{start}} < t < T_{\text{end}} \\ \lambda_{\text{hf}}^{\max}, & t \geq T_{\text{end}} \end{cases} \quad (11)$$

Adopting a coarse-to-fine trajectory, this schedule prioritizes low-frequency structures to stabilize early geometric convergence, thereby mitigating the impact of unstable high-frequency gradients while incorporating sharp edge priors in later stages for detailed reconstruction.

The total training objective integrates photometric supervision, structural constraints, and teacher-guided high-frequency alignment:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pix}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{ssim}} + \lambda_{\text{hf}}(t)\mathcal{L}_{\text{hf}} + \mathcal{R}_{\text{cal}} + \mathcal{R}_{\text{tone}}, \tag{12}$$

where $\mathcal{L}_{pix}$ denotes a photometric consistency loss, $\mathcal{L}_{ssim}$ is the SSIM term, and $\mathcal{L}_{hf}$ is the confidence-masked high-frequency residual loss in Eq. (10). $\mathcal{R}_{cal}$ and $\mathcal{R}_{tone}$ are regularizers for PhysIR-Cal and IRTone-Adapter, respectively, and curriculum learning is realized through the time-varying weight $\lambda_{hf}(t)$.

4 Experiments

4.1 Experimental Settings

Datasets. Evaluation is conducted on two challenging real-world TIR datasets. 1) **TI-NSD** [3] dataset comprises 20 scenes, including 7 indoor, 7 outdoor, and 6 UAV scenarios, covering materials with diverse emissivity and complex heat transfer effects. Following a standard split, every eighth frame is held out for testing, with a native resolution of 720×480. 2) **ThermoScenes** [9] dataset contains 10 scenes (6 indoor and 4 outdoor), spanning indoor objects with steep temperature gradients and outdoor building facades affected by atmospheric transmittance; 1/8 of views are reserved for testing to assess novel-view generalization within each scene, with a native resolution of 480×640. For both datasets, LR inputs are formed through $4\times$ bicubic downsampling of the original images.

Evaluation Metrics. Visual fidelity is quantified using standard PSNR [11], SSIM [30], and LPIPS [38]. To assess temperature fidelity, we report the global Mean Absolute Error (MAE) on the valid pixel set and the Region-of-Interest MAE (MAE_{ROI}) for thermal targets. Following [24], heated foreground objects are identified via Otsu's thresholding. To mitigate the interference of extreme outliers, an outlier rejection scheme is applied where pixels with prediction errors exceeding $T_{\text{err}} = 10°\text{C}$ are excluded from the calculation.

Baselines. Comparisons are conducted against three representative groups under matched settings: (i) interpolation followed by 3DGS, including Bicubic+3DGS and Lanczos+3DGS, as well as models trained directly on LR inputs with $4\times$ rendering scales, namely 3DGS($4\times$) [14], Mip-Splatting($4\times$) [36], Scaffold-GS($4\times$) [20], and Thermal3DGS($4\times$); (ii) cascaded pipeline via 2D super-resolution followed by 3DGS, instantiated as DifIISR+3DGS [16]; and (iii) a recent visible 3D super-resolution baseline, SRGS [6].

Implementation Details. Our model is implemented in PyTorch and optimized for 30,000 iterations on an NVIDIA RTX 3080. We follow the standard 3DGS optimization schedule for learning rates and densification. The frequency-aware curriculum is realized via cosine ramp-ups for the $\mathcal{L}_{\text{ssim}}$ (10%-40%), $\mathcal{L}_{\text{hf}}$ (65%-95%), and late-stage boosting terms (55%-95%). Regularization weights λ_{cal} and λ_{tone} are set to 5×10^{-4} and 10^{-3}, respectively. A pre-trained SwinIR-S

Table 1: Quantitative comparison on the TI-NSD dataset. All methods are evaluated under the LR input setting. The suffix (4×) denotes direct rendering at 4× resolution using the LR-trained model. We differentiate the best, second best and third best methods by colors.

Method	Indoor			Outdoor			UAV		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Bicubic + 3DGS	0.9222	31.16	0.3095	0.8466	27.47	0.3311	0.8399	29.26	0.3127
Lanczos + 3DGS	0.9237	31.04	0.3085	0.8375	26.94	0.3360	0.8444	29.39	0.3069
DifIISR + 3DGS	0.9278	30.45	0.3097	0.8453	26.77	0.3235	0.8449	28.96	0.2899
3DGS (4×)	0.9060	28.23	0.3147	0.7825	21.86	0.3491	0.7555	23.37	0.3562
Thermal3DGS (4×)	0.9154	30.15	0.3117	0.7852	23.36	0.3418	0.7287	23.40	0.3733
Mip-Splatting (4×)	0.9232	30.63	0.3020	0.8280	25.37	0.3304	0.8582	29.92	0.2786
Scaffold-GS (4×)	0.9145	28.70	0.3077	0.7998	22.05	0.3331	0.7405	21.64	0.3612
SRGS	0.9312	32.09	0.2765	0.8701	27.38	0.2790	0.9078	31.78	0.2014
Ours	0.9455	34.81	0.2597	0.9066	29.94	0.2336	0.9428	34.24	0.1453

serves as the teacher network, with pseudo-labels and confidence masks precomputed to minimize runtime memory overhead.

4.2 Comparison results

Results on TI-NSD dataset. Table 1 summarizes the quantitative comparisons on the TI-NSD dataset. Under the low-resolution protocol, our proposed model consistently achieves superior performance across all subsets (Indoor, Outdoor, and UAV). Naive 4× rendering of standard 3DGS methods (e.g., 3DGS and Scaffold-GS) and direct upsampling baselines exhibit significant degradation; for instance, 3DGS yields only 21.86 dB PSNR in Outdoor scenarios. This failure stems from the inherent properties of infrared imaging: thermal diffusion attenuates high-frequency geometric cues, while FPN and NUR artifacts introduce structured noise that misleads standard 3DGS optimization. While cascaded 2D-prior pipelines (e.g., DifIISR+3DGS) enhance fidelity, they lack inherent multi-view consistency, which restricts the accuracy of the reconstructed 3D geometry. Even SRGS, a state-of-the-art 3D visiable super-resolution baseline, shows performance saturation on TIR data. In contrast, our model outperforms SRGS by 2.7 dB and 2.6 dB PSNR in Indoor and Outdoor scenes, respectively, with notably lower LPIPS. We attribute these gains to our physics-aware design: PhysIR-Cal decouples scene radiance from hardware-induced artifacts, mitigating SNR inversion and enabling the model to recover robust geometry from cross-view consistent structures.

Fig. 2 presents visual comparisons for novel-view synthesis. Relative to competing methods, the proposed method better preserves thin structures and delineates geometric boundaries while suppressing hardware-induced artifacts. This behavior aligns with the fact that in TIR imaging, thermal diffusion blurs physical boundaries into low-frequency transitions, while sensor artifacts manifest

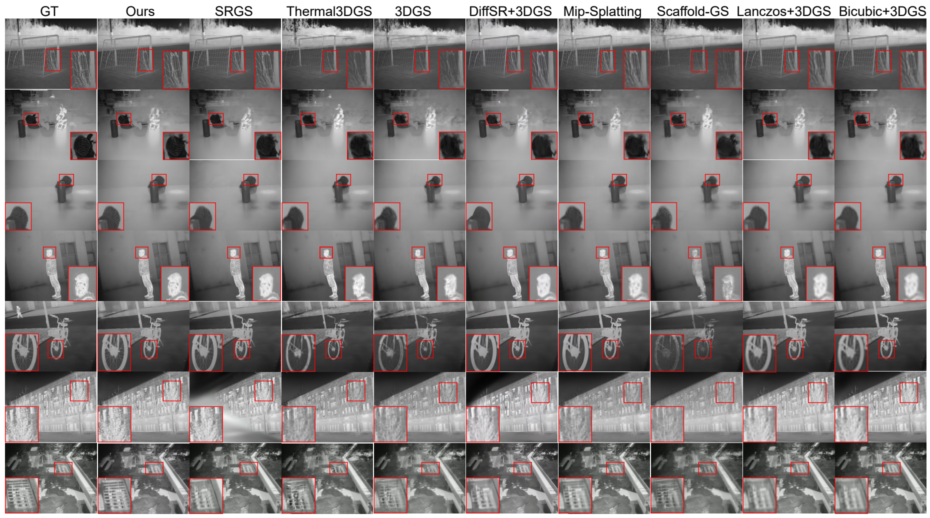

Fig. 2: Qualitative results of $4\times$ **super-resolution novel view synthesis on the TI-NSD dataset.** Our method suppresses floating artifacts and recovers sharper high-frequency geometry. Red boxes highlight zoomed details.

as structured high-frequency textures. Consequently, baseline methods are frequently misled by these spurious signals, leading to pervasive floaters and local geometric collapse. This degradation is particularly evident in delicate structures, such as the soccer net (row 1) and the bicycle hub (row 5). standard 3DGS and its variants (e.g., Mip-Splatting and Scaffold-GS) tend to break connectivity and accumulate irregular edge blobs. While DifIISR-based smoothing reduces noise, it erases physically meaningful details; similarly, SRGS partially restores coarse silhouettes, but without explicit decoupling of IR-specific noise, the renderings remain low-contrast and visually inconsistent.

In contrast, our framework reconstructs fine-scale geometry with fewer floaters. Consequently, the rendering results retain sharper, coherent high-frequency content in challenging areas (e.g., branch boundaries in row 6 and the standing person contour in row 4), approaching the ground-truth appearance.

Results on ThermoScenes dataset. To evaluate model robustness across wide temperature spans and extreme high-contrast scenarios, we conduct experiments on the ThermoScenes dataset. Given that this dataset provides thermal radiation mapping, we focus on the model's fidelity in recovering thermodynamic states and its resilience to sensor noise. Table 2 and Fig. 3 summarize the corresponding quantitative and qualitative results on ThermoScenes dataset.

Quantitative Insights on Smoothness Bias. During this analysis, we observe a classic yet noteworthy 'smoothness bias' phenomenon. Outdoor scenes in thermal imaging typically comprise large areas of flat thermal backgrounds (e.g.,

Table 2: Quantitative comparison on the ThermoScenes dataset. All methods are evaluated under the LR input setting. The suffix (4×) denotes direct rendering at 4× resolution using the LR-trained model. We differentiate the best, second best and third best methods by colors.

Method	Indoor					Outdoor				
	SSIM↑	PSNR↑	LPIPS↓	MAE↓	MAE_{roi} ↓	SSIM↑	PSNR↑	LPIPS↓	MAE↓	MAE_{roi} ↓
Bicubic + 3DGS	0.9650	31.19	0.0415	0.3536	1.6318	0.9663	30.24	0.1577	0.6166	0.5411
Lanczos + 3DGS	0.9665	30.73	0.0414	0.3574	1.7223	0.9686	30.54	0.1555	0.7988	0.7441
DifIISR + 3DGS	0.9486	28.95	0.0530	0.7947	1.9993	0.9617	29.21	0.1636	1.1811	1.1606
3DGS (4×)	0.9612	29.23	0.0556	0.3581	2.0971	0.9238	25.71	0.2164	0.8346	0.7380
Thermal3DGS (4×)	0.9122	27.17	0.0504	0.8558	2.1585	0.9201	23.92	0.2205	1.8974	2.0263
Mip-Splatting (4×)	0.9662	30.59	0.0476	0.3685	1.8094	0.9585	29.83	0.1659	0.7149	0.6316
Scaffold-GS (4×)	0.9656	28.93	0.0545	0.3219	2.1495	0.9315	26.45	0.1998	0.6888	0.6311
SRGS	0.9537	29.84	0.0454	0.4048	2.1793	0.9679	30.76	0.1472	0.6619	0.5969
Ours	0.9672	31.32	0.0438	0.3368	1.7903	0.9662	30.40	0.1546	0.7957	0.7329

sky, cold walls) alongside minute heat sources with extreme local contrast. Consequently, traditional low-pass filtering and 2D interpolation techniques (e.g., Bicubic+3DGS, Lanczos+3DGS) tend to produce overly smoothed renderings.

While this conservative smoothing strategy mathematically circumvents global pixel-wise penalties, yielding deceptively competitive PSNR and MAE scores, it does so at the catastrophic expense of critical high-frequency 3D geometry. For instance, as demonstrated in Fig. 3 (row 2), although Bicubic+3DGS achieves higher PSNR, it completely blurs the critical structural boundaries of the building. In contrast, our proposed method resists this texture degeneration and preserves true physical geometric fidelity. It achieves highly competitive structural similarity (SSIM) and substantially outperforms existing 3D representations (e.g., Scaffold-GS, Mip-Splatting, and Thermal3DGS) across comprehensive metrics, particularly in LPIPS and region-of-interest MAE (MAE_{roi}). By effectively decoupling high-frequency thermal structures from background noise, our approach bypasses the smoothness bias, preserving true physical geometric fidelity rather than defaulting to naive texture blurring.

Qualitative Analysis of Failure Cases. The visual comparisons in Fig. 3 and Fig. 4 highlight distinct failure modes of baseline methods under severe TIR degradations. Simple super-resolution pipelines (e.g., Bicubic, DifIISR+3DGS) struggle to distinguish structural boundaries between window frames and walls, erasing high-frequency content and yielding over-smoothed 3D geometry (rows 1-2). Conversely, standard 3DGS variants overfit structured FPN from low-resolution inputs, leading to agglomerated artifacts and local geometric dropouts near intense heat sources. Furthermore, as demonstrated across diverse scenes (Fig. 4), although the 3D super-resolution baseline SRGS succeeds on visible spectra, it lacks explicit IR-specific modeling. Consequently, it exhibits edge ghosting and detail attenuation on complex targets (e.g., kettles and melting ice cups), failing to reconstruct thin, intricate structures.

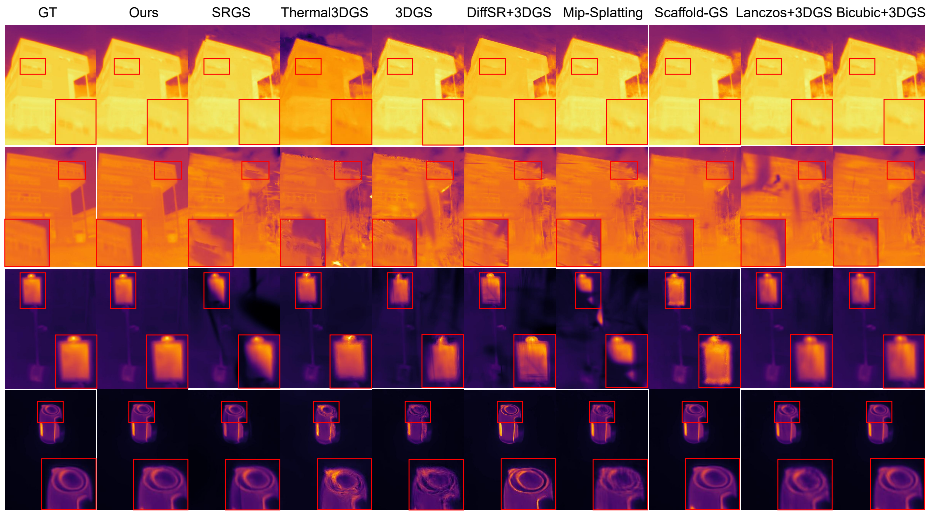

Fig. 3: Qualitative results of $4\times$ **super-resolution novel view synthesis on the ThermoScenes dataset.** Our method successfully suppresses floating artifacts and recovers sharper high-frequency geometry. Red boxes highlight zoomed details.

In contrast, our method superiorly preserves high-frequency details and structural boundaries under extreme thermal contrast. PhysIR-Cal decouples NUR responses from scene radiance to mitigate floater artifacts, while the IRTone-Adapter provides intensity-conditioned modulation to uncover subtle temperature gradients. Driven by frequency-aware curriculum learning, the reconstructed 3D geometry aligns with thermodynamic boundaries, closely approximating the ground-truth appearance even for fine-grained details.

Efficiency Analysis. As seen in Table 3, SupIR-GS adds moderate training overhead from calibration and curriculum modules, but the teacher is discarded during inference. It keeps real-time rendering while achieving the lowest peak memory and highest FPS among compared methods.

Table 3: Efficiency comparison under the same setting: RTX 3080, 30K iterations, and $4\times$ HR rendering.

Method	Time(min)↓	Mem(MB)↓	FPS↑
SRGS	10.49	3484	313.75
3DGS	**10.39**	3580	276.36
Ours	16.42	**2792**	**381.10**

4.3 Ablation Study

Ablation studies were conducted on TI-NSD dataset. The baseline follows a standard 3D-SR pipeline, similar to SRGS. Detailed results are reported in Table 4.

Baseline Limitations. Optimizing standard 3D-SR pipeline on noisy LR inputs underperforms (Row 1). This confirms their inability to handle TIR-specific signal-to-noise inversion, where hardware-induced high-frequency patterns dominate optimization over low-frequency physical boundaries.

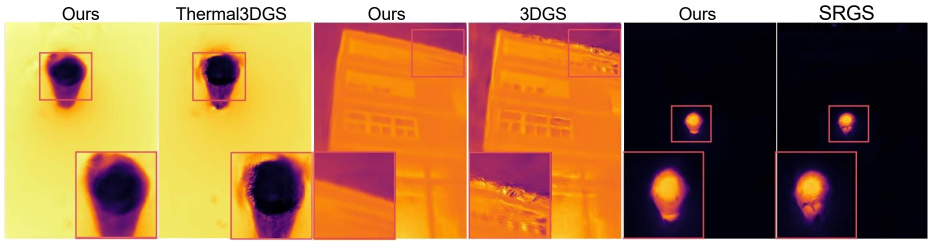

Fig. 4: Visual comparisons of $4\times$ super-resolution. Our method suppresses floating artifacts and over-smoothing, recovering sharper high-frequency geometry. Red boxes indicate zoomed regions.

Table 4: Ablation study of core components. We progressively integrate PhysIR-Cal, IRTone-Adapter, and Frequency-Aware into the baseline. Results show that these components are not only individually effective but also exhibit strong physical synergy.

	Module Configuration			Indoor			Outdoor			UAV		
	PhysIR-Cal	IRTone-Adapter	Frequency-Aware	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
Baseline	×	×	×	0.9312	32.086	0.2765	0.8701	27.381	0.2790	0.8888	30.291	0.2299
	×	✓	×	0.9400	31.700	0.2920	0.8655	27.202	0.2836	0.8988	30.860	0.2087
	×	×	✓	0.9354	32.281	0.2662	0.8727	27.263	0.2680	0.9121	31.550	0.1868
	×	✓	✓	0.9401	32.936	0.2638	0.8925	28.514	0.2491	0.9231	32.203	0.1745
	✓	×	×	0.9420	33.692	0.2837	0.8795	28.449	0.2681	0.9118	32.218	0.1928
	✓	×	✓	0.9453	34.276	0.2788	0.9031	29.800	0.2377	0.9305	33.142	0.1668
Ours	✓	✓	✓	**0.9455**	**34.806**	**0.2597**	**0.9066**	**29.935**	**0.2336**	**0.9320**	**33.326**	**0.1653**

Unguided Enhancement. Integrating the IRTone-Adapter in isolation (Row 2 vs. 1) degrades performance, with Indoor PSNR dropping by 0.39 dB. This shows uncalibrated contrast stretching amplifies high-frequency sensor artifacts into spurious geometry, underscoring the need for prior physical calibration.

Physical Decoupling. Introducing PhysIR-Cal alone (Row 5 vs. 1) yields substantial gains across all subsets, increasing PSNR by 1.61 dB on Indoor and 1.93 dB on UAV. This demonstrates that decoupling sensor noise from the latent thermal radiance field is a prerequisite for forming a reliable geometric scaffold.

Prior Dependency. Applying Frequency-Aware constraints alone (Row 3 vs. 1) yields marginal or negative gains, such as a 0.12 dB PSNR drop on Outdoor. Under structured noise, this induces gradient conflicts, as the teacher prior favors edges while reconstruction losses encourage fitting hardware patterns.

Progressive Module Synergy. A consistent dependency emerges where the high-frequency prior becomes effective only after PhysIR-Cal establishes a denoised scaffold (Row 6). Further enabling the IRTone-Adapter allows local tone modulation to amplify genuine thermal contrast without propagating artifacts, culminating in total PSNR gains of 2.72 dB, 2.55 dB, and 3.04 dB across the three subsets. Together, the three components align with a physically consistent sequence: denoise first, enhance next, then refine high-frequency detail.

Teacher Prior Reliability. Fig. 5 verifies that the RGB-pretrained SwinIR-S teacher does not simply induce 2D replication. Its enhanced TIR details are used as confidence-filtered high-frequency cues only in the late curriculum stage,

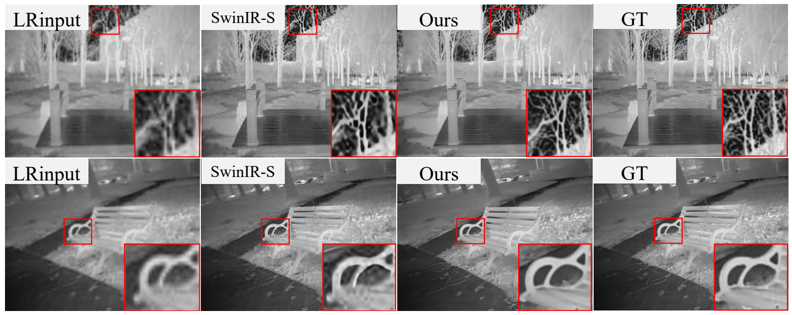

Fig. 5: SwinIR teacher prior visualization. Comparison of LR input, SwinIR-S, SupIR-GS, and GT. Zoomed regions show that SupIR-GS recovers more GT-consistent thermal boundaries than the raw teacher output.

while the 3DGS reconstruction is still constrained by multi-view consistency and remains closer to GT than the raw teacher output.

5 Conclusion

In this paper, we address the challenge of high fidelity 3D thermal infrared reconstruction from low resolution observations. We identify signal-to-noise inversion as a critical bottleneck, where intrinsic thermal diffusion and sensor non-uniformity cause standard optimization to overfit hardware noise while blurring geometric boundaries. We propose SupIR-GS, an imaging calibration inspired framework bridging the gap between latent radiance and degraded observations. At its core, PhysIR-Cal decouples radiance from sensor artifacts via a differentiable layer, while IRTone-Adapter employs local affine residual modulation to surface latent gradients. Optimization is further stabilized by frequency aware curriculum learning, guiding the trajectory from coarse scaffolds to fine grained details. Benchmarks show that SupIR-GS suppresses spurious floaters and recovers sharp boundaries, outperforming baselines in synthesis quality and temperature fidelity. By integrating physics informed priors with explicit 3D representations, our work provides a robust solution for precise thermal scene perception.

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 China Postdoctoral Science Foundation (General Program, Grant No.2024M760716) 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. Chen, L., Tang, R., Anisetti, M., Yang, X.: A lightweight iterative error reconstruction network for infrared image super-resolution in smart grid. Sustainable Cities and Society **66**, 102520 (2021)
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. Chudasama, V., Patel, H., Prajapati, K., Upla, K.P., Ramachandra, R., Raja, K., Busch, C.: Therisurnet-a computationally efficient thermal image super-resolution network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 86–87 (2020)
5. Fan, Z., Wang, K., Wen, K., Zhu, Z., Xu, D., Wang, Z.: Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. Advances in neural information processing systems **37**, 140138–140158 (2024)
6. Feng, X., He, Y., Wang, Y., Yang, Y., Li, W., Chen, Y., Kuang, Z., Fan, J., Jun, Y., et al.: Srgs: Super-resolution 3d gaussian splatting. arXiv preprint arXiv:2404.10318 (2024)
7. Fraternali, P., Morandini, L., Motta, R.: Enhancing search and rescue missions with uav thermal video tracking. Remote Sensing **17**(17), 3032 (2025)
8. Guédon, A., Lepetit, V.: Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5354–5363 (2024)
9. 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)
10. Hu, L., Hu, L., Chen, M.: Edge-enhanced infrared image super-resolution reconstruction model under transformer. Scientific Reports **14**(1), 15585 (2024)
11. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of psnr in image/video quality assessment. Electronics letters **44**(13), 800–801 (2008)
12. 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)
13. Kansal, P., Nathan, S.: Dual-input frequency-aware network for high-quality thermal image super-resolution. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 4610–4620 (2025)
14. 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)
15. Lee, B., Lee, H., Sun, X., Ali, U., Park, E.: Deblurring 3d gaussian splatting. In: European Conference on Computer Vision. pp. 127–143. Springer (2024)
16. Li, X., Wang, Z., Zou, Y., Chen, Z., Ma, J., Jiang, Z., Ma, L., Liu, J.: Difiisr: A diffusion model with gradient guidance for infrared image super-resolution. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 7534–7544 (2025)

17. Liang, S., Song, K., Zhao, W., Li, S., Yan, Y.: Dasr: Dual-attention transformer for infrared image super-resolution. Infrared Physics & Technology **133**, 104837 (2023)
18. 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)
19. 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)
20. 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)
21. 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)
22. 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)
23. Munir, F., Azam, S., Rafique, M.A., Sheri, A.M., Jeon, M., Pedrycz, W.: Exploring thermal images for object detection in underexposure regions for autonomous driving. Applied soft computing **121**, 108793 (2022)
24. Otsu, N., et al.: A threshold selection method from gray-level histograms. Automatica **11**(285-296) (1979)
25. Patel, H.: Thermal image super-resolution challenge-pbvs 2020. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2020)
26. Rangel, J., Soldan, S., Kroll, A.: 3d thermal imaging: Fusion of thermography and depth cameras. In: International Conference on Quantitative InfraRed Thermography. vol. 3 (2014)
27. Shen, Y., Ceylan, D., Guerrero, P., Xu, Z., Mitra, N.J., Wang, S., Frühstück, A.: Supergaussian: Repurposing video models for 3d super resolution. In: European Conference on Computer Vision. pp. 215–233. Springer (2024)
28. Tong, Z., Cheng, L., Xie, S., Kersemans, M.: A flexible deep learning framework for thermographic inspection of composites. NDT & E International **139**, 102926 (2023)
29. Wang, C., Wu, X., Guo, Y.C., Zhang, S.H., Tai, Y.W., Hu, S.M.: Nerf-sr: High quality neural radiance fields using supersampling. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6445–6454 (2022)
30. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
31. 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)
32. Yoon, Y., Jang, H.K., Yoon, K.J.: Gmt: Enhancing generalizable neural rendering via geometry-driven multi-reference texture transfer. In: European Conference on Computer Vision. pp. 274–292. Springer (2024)
33. Yoon, Y., Yoon, K.J.: Cross-guided optimization of radiance fields with multi-view image super-resolution for high-resolution novel view synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12428–12438 (2023)

34. Yoon, Y., Yoon, K.J.: Resplat: Degradation-agnostic feed-forward gaussian splatting via self-guided residual diffusion. In: The Fourteenth International Conference on Learning Representations (2026)
35. Yu, X., Zhu, H., He, T., Chen, Z.: Gaussiansr: 3d gaussian super-resolution with 2d diffusion priors. arXiv preprint arXiv:2406.10111 (2024)
36. 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)
37. Zeng, H., Bai, Y., Fu, Y.: Arbitrary-scale 3d gaussian super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 12304–12312 (2026)
38. 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)
39. Zhao, S., Fang, Z., Wen, S.: 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). pp. 289–294. IEEE (2017)