Thermal is Always Wild: Characterizing and Addressing Challenges in Thermal-Only Novel View Synthesis

M. Kerem Aydin[1] Vishwanath Saragadam[2] Emma Alexander[1]
[1]Northwestern University [2]University of California, Riverside
https://nubivlab.github.io/wild_thermal

Abstract

Thermal cameras provide reliable visibility in darkness and adverse conditions, but thermal imagery remains significantly harder to use for novel view synthesis (NVS) than visible-light images. This difficulty stems primarily from two characteristics of affordable thermal sensors. First, thermal images have extremely low dynamic range, which weakens appearance cues and limits the gradients available for optimization. Second, thermal data exhibit rapid frame-to-frame photometric fluctuations together with slow radiometric drift, both of which destabilize correspondence estimation and create high-frequency floater artifacts during view synthesis, particularly when no RGB guidance (beyond camera pose) is available. Guided by these observations, we introduce a lightweight preprocessing and splatting pipeline that expands usable dynamic range and stabilizes per-frame photometry. Our approach achieves state-of-the-art performance across thermal-only NVS benchmarks, without requiring any dataset-specific tuning.

1. Introduction

Thermal cameras capture long-wavelength radiation emitted and reflected by surfaces, revealing structure that is often invisible to RGB sensors. Thermal cameras function reliably under conditions that challenge visible-light imaging; such as darkness, fog, or smoke, and expose information related to temperature and material properties. These characteristics make thermal imaging a valuable sensing modality for applications including autonomous driving [2, 5], environmental monitoring [33], infrastructure inspection [24], robotics [6] and search-and-rescue [43]. Extending these capabilities to 3D perception through novel view synthesis (NVS) would enable reliable scene reconstruction in settings where visible-light cameras fail.

Novel view synthesis (NVS) enables reconstructing scene geometry and appearance from posed image collections and has become a reliable 3D perception tool for RGB imagery, supporting applications in robotics [34], autonomous driving [13], and AR/VR [18]. Its effectiveness stems from the rich texture, stable photometry, and consistent multiview observations typically available in visible-light data. These assumptions, however, do not hold for thermal imagery, making NVS substantially more challenging despite its benefits in low-visibility settings.

Thermal images lack the chromatic and textural cues that support reliable correspondence in RGB and exhibit sensor-induced inconsistencies that disrupt geometric and photometric agreement across views. These violations of multiview consistency make direct translation of RGB-based NVS pipelines unstable and often lead to geometric errors or radiometric artifacts. Consequently, most prior work relies on paired RGB–thermal inputs [12, 20, 22, 29, 38, 46], using RGB to recover structure while treating thermal measurements as an auxiliary channel. However, this dependence on visible-light imagery limits applicability in the very conditions where thermal sensing is most valuable, motivating the need for reliable thermal-only NVS. Fig. 1 highlights the challenges in thermal data and illustrates our SOTA NVS performance.

This paper provides a rigorous framework for evaluating and improving thermal-only NVS by examining how thermal-specific degradations appear in real data. To understand how these issues manifest in practice, we examine several public multiview thermal datasets and document their shared characteristics: low dynamic range, photometric fluctuations, slow radiometric drift, and limited texture across views. Although the severity of these effects varies by dataset, they consistently reduce the stability of multiview correspondence and motivate the need for preprocessing steps that normalize thermal observations before reconstruction. Guided by these observations, we develop a Gaussian-splatting pipeline tailored to the characteristics of thermal imagery. Our approach integrates a lightweight preprocessing module that expands usable dynamic range and stabilizes per-frame photometry, followed by a splatting stage adapted to operate reliably under the reduced dynamic range, texture and radiometric variability of thermal

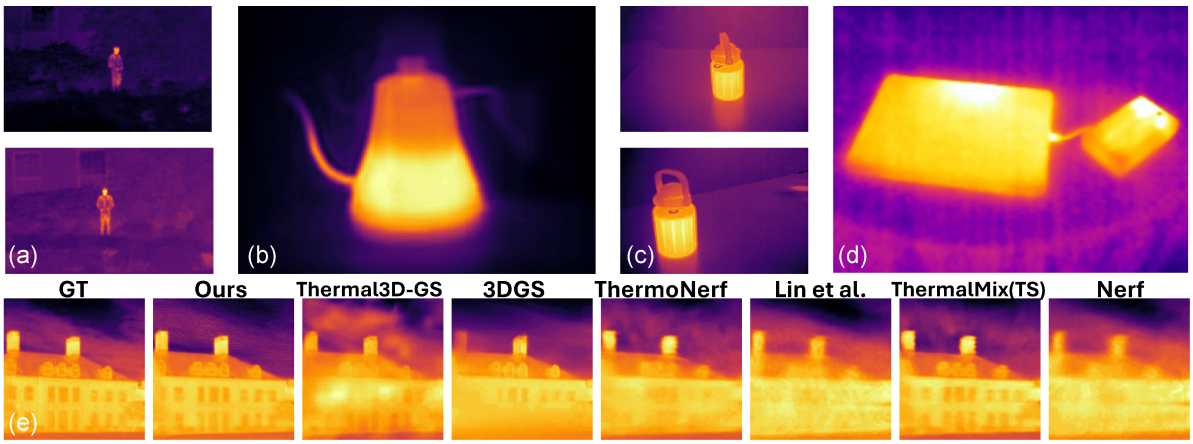

Figure 1. **Our method overcomes significant challenges in thermal images.** Thermal data contain limited texture and lack multispectral cues, making correspondence estimation harder than in RGB. They also exhibit sensor-specific degradations, including (a) frame-to-frame photometric inconsistency from sensor heating, (b) softened transitions between hot and cold regions characteristic of microbolometer sensors, (c) vignetting that produces viewpoint-dependent attenuation, and (d) fixed-pattern noise visible as structured artifacts. Our method explicitly stabilizes the photometry in (a), while the effects in (b–d) are mitigated in our SOTA reconstructions (e) through multiview consistency enabled by a novel embedding approach for learned appearance modeling.

data. Together, these components enable high-fidelity NVS from purely thermal inputs and improve performance across diverse datasets without dataset-specific tuning.

2. Related Work

2.1. Novel View Synthesis

Novel view synthesis (NVS) has advanced rapidly through representations that recover continuous scene structure from posed image collections. Neural Radiance Fields (NeRF) made a major breakthrough by optimizing a volumetric radiance field to reproduce observed RGB views with high fidelity [25]. Subsequent work improved robustness and efficiency through antialiasing [1] and fast multiresolution encodings [27]. Explicit approaches further pushed efficiency by replacing neural volumetric fields with sparse voxel grids [9] or with the anisotropic primitives used in 3D Gaussian Splatting [16]. Together, these advances illustrate the shift from slow neural radiance fields toward fast, explicit primitives that support high-quality NVS.

Standard NVS pipelines degrade when appearance varies across views, leading to drifting color estimates, incorrect density inference, and unstable geometry. Several methods address these issues through appearance modeling, including per-frame embeddings for "in-the-wild" scenarios that adapt to illumination or style changes [23], or through joint pose refinement [19]. Other approaches target extreme low-light conditions by training directly on noisy raw measurements [26]. Recent extensions adapt Gaussian-splatting representations to uncontrolled photo collections by incorporating robust initialization, exposure normalization, or appearance conditioning [7, 17, 37, 39, 44]. Collectively, these works highlight the need for specialized techniques for NVS under appearance shifts, low-light noise, or pose uncertainty. These challenges are amplified in thermal imaging, where low contrast, radiometric drift, and imprecise poses further destabilize reconstruction. In this work, we extend the in-the-wild paradigm to thermal NVS, adapting appearance modeling to address these domain-specific challenges.

2.2. Thermal Imaging

Thermal cameras, and in particular the affordable microbolometer sensor-based images, suffer from strong artifacts that make NVS particularly challenging. First, thermal images contain fixed pattern noise due to fabrication imperfections [14, 31]. Numerous hardware [28] and algorithm-based [10, 11, 14, 21, 32, 36] solutions exist, but they often only partially address the fixed pattern noise. Second, thermal cameras suffer from drifts due to internal heating, causing poor radiometric consistency [36]. As we will see later, this has a debilitating effect on the quality of NVS images. Finally, thermal images suffer from ghosting due to thermal inertia [30], and lack texture, implying both pose estimation, and training the radiance fields is challenging.

2.3. Thermal+RGB NVS

RGB–thermal NVS methods leverage RGB imagery to compensate for the weak texture, radiometric drift, and pose instability characteristic of thermal sensors. In [29], the authors evaluate strategies for integrating thermal inputs into radiance-field pipelines, showing that explicitly adding a second thermal branch yields sharper reconstructions than fine-tuning or single-branch designs. Lin et al. [20] takes a

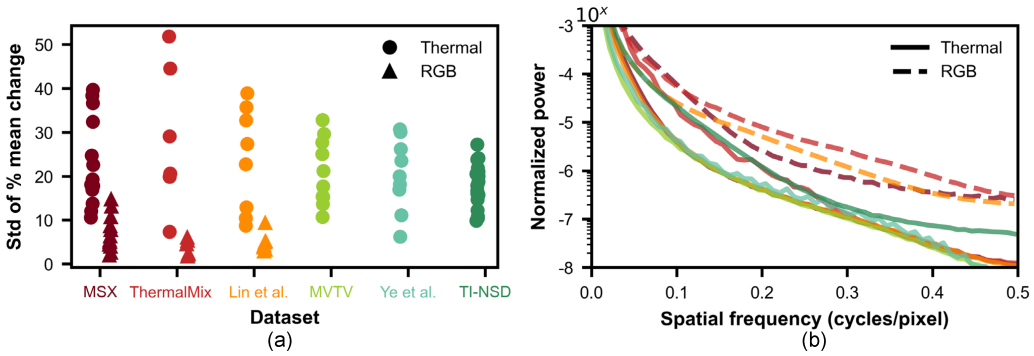

Figure 2. **Dataset-level radiometric and spatial-frequency characteristics.** (**a**) Standard deviation of the relative mean-intensity change ΔI_t across scenes. Thermal sequences show substantially larger radiometric fluctuations than RGB, with variation levels differing noticeably across datasets. (**b**) Radially averaged power spectra, averaged per dataset. Thermal frames consistently exhibit reduced high-frequency energy compared to RGB; among thermal datasets, TI-NSD shows comparatively stronger high-frequency content.

two-model approach, optimizing separate RGB and thermal NeRFs while enforcing cross-modal consistency through an L_1 density regularizer, allowing RGB to guide the geometry while preserving modality-specific appearance. ThermoNeRF [12] applies a similar paired-modality formulation to architectural scenes, using aligned RGB–thermal images to learn density while predicting color and temperature through separate networks. Additional work uses thermal measurements to improve reconstruction in low-light conditions [38, 46] or adapts Gaussian splatting to thermal inputs [22], and other approaches incorporate thermal cues to stabilize RGB NVS in smoke-filled environments[15]. Together, these methods demonstrate that RGB supervision significantly improves pose stability, geometry quality, and radiometric consistency for thermal data, but they require paired multimodal capture, motivating the need for thermal-only NVS methods.

2.4. Thermal-Only NVS

Thermal-only NVS aims to reconstruct geometry and appearance directly from thermal imagery, but prior work shows that weak texture, low dynamic range, and radiometric instability cause standard NeRF or 3DGS pipelines to struggle on thermal-only data [20, 29]. Here we consider the problem of view synthesis only, assuming that camera poses are known for all methods. In practice, estimating camera poses from thermal images in also difficult, requiring RGB images or other sensors (e.g., IMUs). Thermal-NeRF learns radiance fields from a single infrared camera by introducing thermal mapping that normalizes intensity responses to 0-255 range and a structural patch constraint that stabilizes training in low-texture regions [41]. Thermal3D-GS adapts Gaussian splatting to thermal data through physics-inspired modeling and temperature-consistency constraints, and demonstrates improved reconstruction on a newly collected thermal multiview dataset [4]. Other thermal-only approaches explore emissive–residual Gaussian decompositions [35], degradation-aware radiance-field optimization for blurry or rolling-shutter thermal inputs [3], thermal radiance prediction with physically motivated rendering [8], and segementation-based preprocessing schemes that enrich thermal signals for pose recovery [45]. Together, these works show that thermal-only NVS is feasible but remains strongly limited by low-dynamic-range, blur, and radiometric instability inherent to thermal sensors.

3. An Analysis of Multiview Thermal Datasets

Multiview thermal datasets display a range of sensor behaviors that differ from those commonly encountered in RGB imagery. Because these characteristics influence how reliably an NVS pipeline can interpret thermal observations, it is important to examine how they appear in real sequences before introducing a reconstruction method. In this section, we analyze several representative datasets to identify recurring properties of thermal multiview capture that are most relevant to downstream NVS performance.

We analyze six publicly available multiview thermal datasets used across recent thermal NVS studies: Lin et al.[20], Ye et al.[41], MVTV[38], MSX[22], ThermalMix[29], and TI-NSD[4]. These datasets differ in sensor type, acquisition protocol, and scene content, spanning both uncooled microbolometers and cooled detectors, indoor and outdoor environments, and static and mobile capture setups. This diversity allows us to separate dataset-specific artifacts from properties that consistently appear across thermal imagery.

We characterize dataset behavior using three diagnostics that capture photometric stability, spatial-frequency content, and effective dynamic range. To assess radiometric stability, we measure the relative change in mean intensity

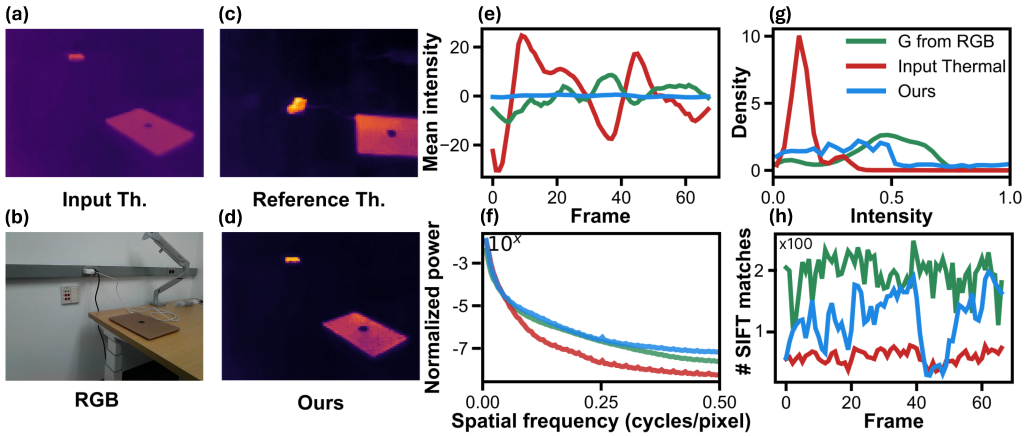

Figure 3. **Photometric stabilization and contrast enhancement** (a) An input thermal frame, with less contrast and texture than the corresponding RGB image (b). Notice the photometric drift when compared to reference frame (c), which shows the same scene at a different time point. (d) Our invertible enhancement improves photometric inconsistency and image contrast. (e) Temporal mean intensity across frames showing reduced radiometric drift after stabilization. (f) Normalized spatial frequency spectrum with enhanced thermal data resembling RGB statistics. (g) Our preprocessing expands the effective dynamic range of thermal images, shown with pixel instensity distributions on a sample frame. (h) Improved structural consistency and stability, illustrated by an elevated number of tracked SIFT features per frame. See Sec. S1 for more examples.

across frames,

$$\Delta I_t = \frac{\mu_t - \bar{\mu}}{\bar{\mu}}, \tag{1}$$

where μ_t is the mean pixel value of frame t and $\bar{\mu}$ is the average intensity across the sequence. Large fluctuations in ΔI_t indicate exposure drift or sensor-heating effects that disrupt brightness constancy and can destabilize correspondence estimation, often producing floater artifacts. Dataset-level trends appear in Fig. 2a, with additional per-scene examples in the supplemental material (Sec. S1.1). To analyze spatial-frequency characteristics, we compute radially averaged power spectra,

$$S_t(f) = \frac{1}{N_f} \sum_{(u,v):\|(u,v)\|\approx f} \left|\mathcal{F}(I_t)(u,v)\right|^2, \tag{2}$$

which describe how image energy is distributed across frequencies. Higher frequencies correlate with texture, sharp edges, and sensor noise. Across datasets, thermal frames exhibit attenuated high-frequency responses due to the microbolometer's smoothing behavior. This reduction in spatial detail weakens geometric and photometric cues that NVS methods typically rely on for stable multiview alignment. Dataset-wide frequency trends are summarized in Fig. 2b, with per-scene examples provided in the supplemental material (Sec. S1.2). Finally, we examine pixel-intensity histograms to evaluate effective dynamic range. Thermal images often occupy a narrow portion of the available intensity space, leading to reduced contrast and weaker optimization gradients. See Fig. 3 for representative histogram behavior with additional examples at Sec. S1.3. Together, these diagnostics highlight the radiometric and contrast-related characteristics that shape thermal multiview data and identify the input properties that most challenge existing NVS pipelines.

In addition to these dominant trends, smaller artifacts such as vignetting and fixed-pattern noise further complicate multiview consistency. The following section introduces a lightweight preprocessing and splatting pipeline informed by these observations.

4. Method

Our approach consists of two components: a contrast–stabilizing preprocessing stage and an adaptation of 3D Gaussian Splatting (3DGS) to the thermal domain. The preprocessing normalizes temporal radiometry and enhances effective contrast, while the splatting stage models thermal appearance using a scalar emission representation conditioned on per-frame and per-Gaussian embeddings (Fig. 4).

4.1. Photometric Stabilization & Enhancement

Thermal sequences exhibit frame-to-frame drift and low effective dynamic range (Sec. 2). We address these effects using our two-step, monotonic transformation composed of sequential histogram alignment and brightness-preserving bi-histogram equalization (BBHE).

Let I_t denote the t-th frame, and let $x \in [0, 1]$ denote a normalized pixel intensity. We maintain an exponentially averaged reference Cumulative Distribution Function

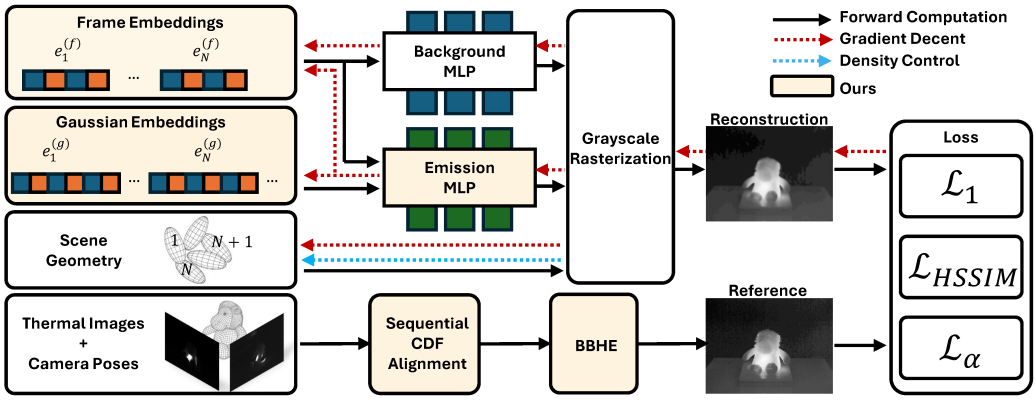

Figure 4. **Our pipeline.** Given thermal frames and camera poses, we first stabilize the inputs to ensure consistent training data across views (bottom). The frames are modeled with a novel combination of per-Gaussian embeddings, which encode spatial appearance, per-frame embeddings, which capture residual temporal artifacts, and a physics-restricted parameter set (grayscale with no spherical harmonics) that stabilizes learning. These components jointly enable consistent thermal reconstruction while preserving fine geometric and intensity details.

(CDF) F_t^*,

$$F_t^*(x) = (1-\alpha)F_{t-1}^*(x) + \alpha F_t(x), \tag{3}$$

where F_t is the CDF of I_t and α controls temporal smoothing. Each frame is stabilized by mapping its distribution to the reference:

$$I_t'(x) = (1-\beta)x + \beta\, F_t^{*-1}(F_t(x)), \tag{4}$$

with $\beta \in [0,1]$ balancing identity and alignment. This step suppresses photometric drift while adapting smoothly to gradual scene changes. We apply BBHE to I_t', splitting the histogram at its mean intensity T_t^μ and equalizing the lower and upper subranges independently:

$$\hat{I}_t(x) = \begin{cases} T_t^L(x), & x \le T_t^\mu, \\ T_t^U(x), & x > T_t^\mu. \end{cases} \tag{5}$$

Because both operations are monotonic and one-to-one, the overall transformation is analytically invertible and preserves a recoverable mapping to the original radiometric scale through a LUT. See Fig. 3 and Sec. S1 for examples of stabilized and contrast-enhanced frames.

4.2. Modeling "Wildness" for Thermal Learning

We adapt 3D Gaussian Splatting to thermal data by simplifying the emission model and incorporating appearance embeddings to handle the residual inconsistencies observed after preprocessing. In thermal imaging, emission is effectively isotropic and single-channel; therefore, instead of predicting RGB colors with spherical harmonics, each 3D Gaussian stores a single scalar emission value. This reduces the complexity of color modeling and aligns the model to the physics of thermal imaging but it also makes the reconstruction more sensitive to frame-dependent fluctuations.

To address these residual inconsistencies, we adopt an embedding-based appearance formulation following the "3DGS-in-the-Wild" literature [7, 17, 37, 39, 44]. Each Gaussian and each frame is assigned a small learnable embedding vector, $\mathbf{e}_i^{(g)}$ and $\mathbf{e}_t^{(f)}$, respectively, and a lightweight MLP maps these embeddings to a scalar emission,

$$c_i(t) = f_\theta\big(\mathbf{e}_i^{(g)}, \mathbf{e}_t^{(f)}\big), \tag{6}$$

which allows the model to absorb smooth frame-dependent variations caused by microbolometer heating, mild vignetting, or fixed-pattern noise without distorting the underlying geometry. These emission values are then used in the usual 3DGS transmittance formulation,

$$\hat{I}_t(\mathbf{r}) = \sum_i T_i\, \alpha_i\, c_i(t), \tag{7}$$

where T_i denotes accumulated transmittance and α_i the opacity of each Gaussian along the ray. During inference, fixing the frame embedding produces temporally stable reconstructions while leaving spatial detail unaffected.

As in recent RGB 3DGS algorithms, we use a background MLP to handle distant regions[37, 40]. The rendered intensity blends foreground and background,

$$\tilde{I}_t(\mathbf{r}) = (1-m(\mathbf{r}))\,\hat{I}_t(\mathbf{r}) + m(\mathbf{r})\, b_\phi(\mathbf{d}, \mathbf{e}_t^{(f)}), \tag{8}$$

where $m(\mathbf{r}) = \exp(-\sum_i \alpha_i)$ is the residual transmittance along the ray.

We train the model with a weighted sum of L1 error, heat-aware Structural Similarity Index Measure (SSIM)[41], and a background regularizer[40]:

$$\mathcal{L} = \lambda_1\, \mathcal{L}_{\text{L1}} + \lambda_2\, \mathcal{L}_{\text{HSSIM}} + \lambda_3\, \mathcal{L}_\alpha. \tag{9}$$

Table 1. **Thermal-only NVS comparison across six multiview datasets.** We report mean PSNR and SSIM across all scenes for six publicly available datasets. Our analysis of dataset difficulty predicts systematic trends in performance across methods. By designing a pipeline that addresses thermal data challenges directly, we deliver SOTA performance on thermal-only NVS.

Decreasing Difficulty

Models	MSX		ThermalMix		MVTV		Lin et al.		Ye et al.		TINSD		Training Time↓
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
Nerf	18. 76	0.62	20.12	**0.71**	19.74	0.72	19.56	0.67	25.08	0.81	26.99	0.89	30+ h
ThermalMix-TS (Instant-NGP)	19.94	0.62	20.67	0.53	18.17	0.70	20.81	0.73	22.34	0.81	25.72	0.91	44m
Lin et al.*	17.31	0.68	16.53	0.70	15.39	0.75	19.73	0.85	20.17	0.83	23.23	0.87	93m
ThermoNerf*	18.49	0.66	18.89	0.72	15.56	0.52	19. 03	0.70	21.84	0.84	27.44	0.92	51m
3DGS	19.87**	0.63	20.37	0.72	15.51**	0.81	20.18	0.74	22.19	0.86	29.04	0.94	**5m**
Thermal3D-GS	22.28	0.64	23.99	**0.77**	22.79**	0.83	23.56	**0.88**	26.54	0.89	31.83	**0.95**	9m
Ours	**23.59**	**0.71**	**24.61**	0.74	**25.13**	**0.84**	**26.37**	0.78	**28.10**	**0.92**	**32.94**	0.94	11m

*Model has its RGB components removed.
**Excludes cases with convergence failure.

The term $\mathcal{L}_{\text{HSSIM}}$ emphasizes thermal contrast and structure, while $\mathcal{L}_{\alpha}$ discourages floaters representing background regions.

5. Results

5.1. Implementation Details

Our system is implemented using the differentiable rasterization backend of GSplat [42], adapted for single-channel thermal imagery. Each Gaussian stores a single emission coefficient without spherical harmonics, and the emission MLP uses three hidden layers of width 128 with ReLU activations followed by a linear output layer. We train with the Adam optimizer and weight decay, using distinct learning rates for geometry, opacity, and appearance parameters. All scenes are rendered at 1080p resolution; for quantitative evaluation, we crop and resize the outputs to match the original sensor resolution for pixel-wise alignment with ground truth. Training runs for 30k iterations on an NVIDIA RTX A6000 GPU without learning-rate scheduling. Gaussian centers are initialized from sparse COLMAP reconstructions, which recover stable but relatively sparse geometry in thermal data.

5.2. Results and Comparison

We compare our method against both general-purpose and thermal-specific neural reconstruction frameworks. As standard references, we include NeRF [25] and 3D Gaussian Splatting (3DGS) [16]. To assess performance against recent thermal pipelines, we evaluate the ThermalMix-TS variant [29], derived from InstantNGP and designed for thermal sequences, and two cross-modal approaches, Lin et al.'s ThermalNeRF [20] and ThermoNeRF [12]. Though the latter two originally rely on RGB–thermal supervision, we disable their RGB branches and cross-modal regularization losses to evaluate thermal-only performance. Finally, we include Thermal3D-GS [4], a thermal-only adaptation of Gaussian splatting that serves as the current state of the art. All methods are trained under matching data splits, iteration counts, and hardware configuration to provide a consistent evaluation. Quantitative results are summarized in Tab. 1, and representative novel-view renders are shown in Figs. 1 and 5.

NeRF and ThermalMix-TS serve as baseline volumetric methods for thermal reconstruction. Across all datasets in Tab. 1, both models exhibit limited performance, with NeRF consistently ranking among the lowest in PSNR and SSIM due to its reliance on high-frequency RGB cues that are largely absent in thermal imagery. As shown in Figs. 1 and 5, NeRF produces blurred surfaces and washed-out edges on the MSX Building and MVTV Mason scenes, and fails to recover background temperature on the Generator example. ThermalMix-TS improves quantitative accuracy slightly, benefiting from faster optimization and better edge localization, yet it remains susceptible to noise and inconsistent global brightness. While its hash-based representation enhances sharpness in small details, it also leads to unstable radiometric behavior across frames, leading to artifacts in video reconstructions. Overall, both methods reveal the inherent limitations of RGB-oriented volumetric frameworks when applied to thermal data.

Both Lin et al. and ThermoNeRF were originally designed as cross-modal methods that jointly leverage RGB and thermal supervision. In their original formulations, the RGB branch primarily drives geometric reconstruction, while the thermal channel acts as an auxiliary modality that paints radiometric information onto the recovered geometry and provides additional regularization. When trained using thermal input alone, this coupling breaks down: both models lose geometric consistency and fail to recover fine-scale texture, producing oversmoothed surfaces and flattened temperature gradients. As shown in Tab. 1, their

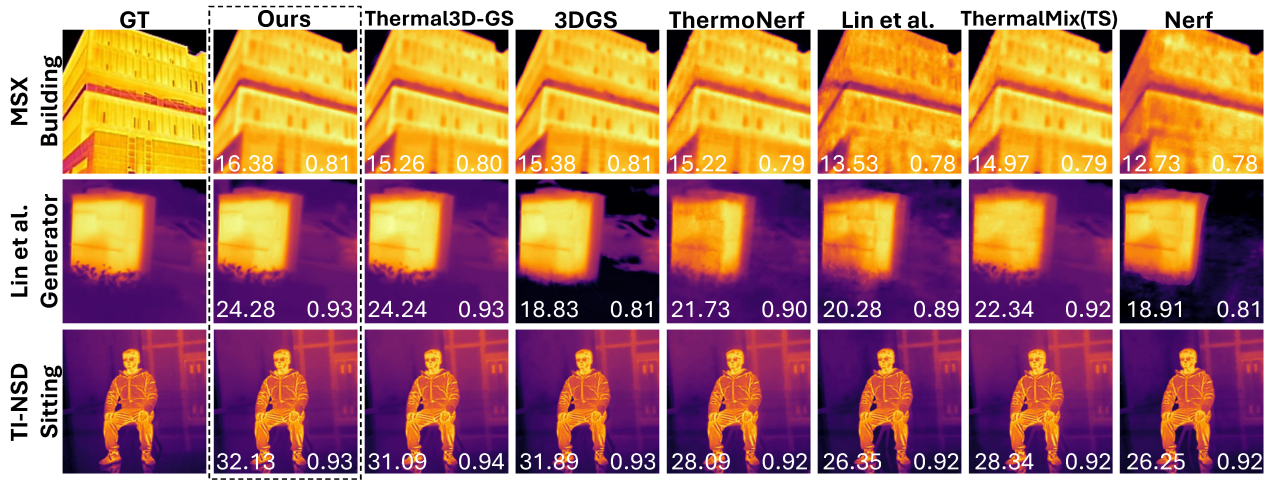

Figure 5. **Comparison to all methods.** Representative novel views, PSNR (bottom left, dB), and SSIM (bottom right) across datasets. Our method yields sharper boundaries and more stable background temperature on challenging scenes (top rows), while maintaining competitive quality on easier scenes (bottom rows).

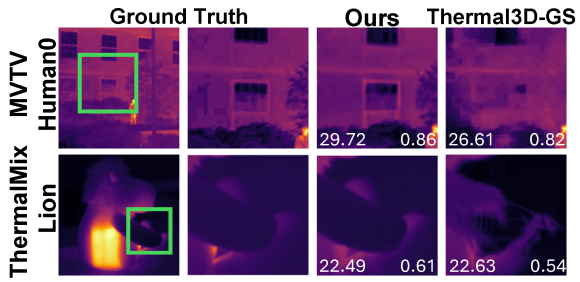

Figure 6. **Comparison to Thermal3D-GS.** Despite competitive metrics (PSNR in dB left, SSIM right), Thermal3D-GS exhibits significant failures in geometry and texture, while our method shows texture like the Human0 window, tree, and smooth wall, and faithfully reconstructs the Lion paw shape.

overall performance drops sharply across all datasets relative to thermal+RGB results previously reported in the literature. Qualitatively (Fig. 5), both methods achieve slightly better geometry than the baseline NeRF but still exhibit blurred edges and weak contrast. In sequences with narrow-baseline training views, correspondence estimation becomes unstable, often leading to ghosting and duplicate surface artifacts in video reconstructions(see generator example in supplementary material). These results demonstrate that cross-modal frameworks lose much of their effectiveness in a purely thermal setting, underscoring the need for reconstruction models designed from the ground up for thermal data.

Among Gaussian-splatting methods, standard 3DGS provides a strong geometry prior but performs inconsistently on thermal data. As reflected in Tab. 1, it achieves reasonable SSIM but suffers from notable PSNR drops on ThermalMix and MVTV, and fails to converge on challenging MSX Iron Ingot/Landscape and MVTV Tree scenes. These stability issues are visible in Fig. 5 generator example, where the background collapses. Thermal3D-GS improves stability through its ATF and TCM modules, yielding better numerical performance and generally sharper reconstructions than 3DGS. However, as shown in Fig. 6, the method fails to capture accurate texture and geometry on scenes where we succeed, and does not converge on the MVTV Tree scene. These failures may be because their ATF module, originally designed to model atmospheric effects (and hence primarily attenuation of intensity), does not model the full variety of photometric inconsistencies that are inherent to thermal images. When the sequence contains unusually bright frames, Thermal3D-GS underfits them, causing those observations to reappear as floating artifacts across a range of novel views. Fig. 7 shows where a single photometrically inconsistent training frame induces floaters that gradually assemble into an entire training viewpoint in Thermal3D-GS, while our method remains stable. Our method avoids these failures not by modeling atmospheric scattering, but by directly addressing the underlying issue: thermal sequences contain radiometric fluctuations that require a representation with higher flexibility than ATF's attenuation-based formulation. By using photometrically stabilized inputs and an embedding-conditioned emission model with greater expressiveness, our approach can represent both bright and dark observations within a consistent radiometric space, preventing floaters and preserving geometry across views. This advantage is consistent across datasets, where we obtain the best or second-best PSNR/SSIM in Tab. 1 while maintaining stable behavior even on the most challenging sequences. These differ-

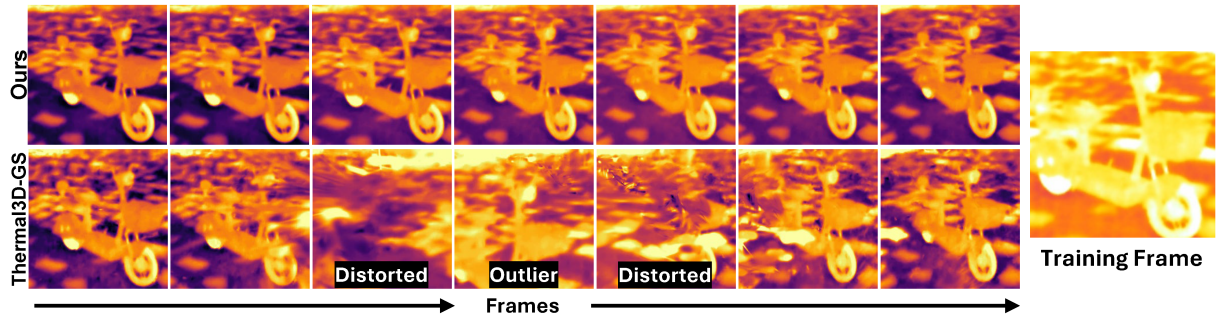

Figure 7. **Photometric consistency promotes high-quality reconstruction.** We render a smooth camera path for our method (top) and Thermal3D-GS (bottom). Thermal3D-GS produces bright floating structures that drift across frames and then assemble into a copy of the photometrically inconsistent training frame (right) when the viewpoint aligns. Our preprocessing and embedding-conditioned emission model handle this outlier without introducing floaters, yielding stable geometry throughout the trajectory.

Table 2. **Ablation study.** We demonstrate that our preprocessing algorithm and "in-the-wild" architecture (3DGS + Emission MLP) independently improve 3DGS performance, and combining them yields superior results. Additional analysis comparing our method to traditional histogram equalization and evaluating each preprocessing step is provided in Tab. S1

Method	MSX- Ebike		T. Mix - Lion		MVTV - Human		Lin et al. - Sink		Ye et al. - Seq.1		TINSD - Sitting		Avg.	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3DGS (Baseline)	20.45	0.86	19.25	0.71	21.21	0.81	20.81	0.74	28.24	0.83	29.51	**0.88**	22.25	0.81
3DGS + Preprocessing	22.79	0.86	24.11	**0.82**	22.01	0.84	23.71	0.78	29.90	0.85	22.42	0.85	23.01	0.83
3DGS + Emission MLP	25.73	0.89	24.17	**0.82**	24.22	0.89	**24.57**	0.83	32.12	0.89	25.98	0.87	24.93	0.87
Ours	**25.97**	**0.92**	**24.25**	0.81	**26.18**	**0.90**	24.27	**0.88**	**33.32**	**0.90**	**30.01**	0.87	**26.14**	**0.88**

ences are even clearer in the supplemental videos, where long camera paths reveal temporal stability and the absence of floaters more clearly than still images. Although our per-scene training time (∼11 min) is slightly longer than 3DGS (5 min) and Thermal3D-GS (9 min), the improvement in reconstruction fidelity and temporal stability represents a favorable quality–efficiency trade-off.

5.3. Ablation Study

As shown in Tab. 2, our preprocessing and emission modeling provide complementary improvements for thermal reconstruction. Preprocessing alone provides modest but consistent gains by stabilizing frame-to-frame radiometric variation, improving average PSNR from 22.25 to 23.01 dB. Our in-the-wild baseline, implemented with an emission MLP and embeddings, models residual temporal artifacts that fixed Gaussian colors cannot capture, raising average PSNR to 24.93 dB. Combining both components produces the best overall performance (26.14 dB / 0.88 SSIM), as preprocessing also introduces stronger gradients that benefit the emission model. Improvements are particularly strong on Human, Ebike, Lion, and Seq.1, while Sitting shows smaller gains due to the already high quality of the baseline reconstruction. A detailed breakdown of preprocessing components and comparison to traditional histogram equalization are provided in Sec. S3.

6. Conclusion

We presented a thermal NVS pipeline designed to address two persistent challenges in thermal imagery: strong frame-to-frame photometric inconsistencies and limited dynamic range. Our approach combines a lightweight, invertible photometric stabilization and contrast enhancement stage with a thermal variant of 3D Gaussian Splatting. The preprocessing aligns each frame to a temporally smooth reference distribution, producing stable, contrast-enhanced inputs for reconstruction. On top of this, we adapt appearance-modeling components from in-the-wild NVS and incorporate a small background-emission MLP tailored to single-channel thermal data. These components allow the model to absorb residual radiometric transients and aberrations. Together, they make thermal NVS feasible without RGB supervision and improve reconstruction stability and fidelity under real-world conditions.

Despite these advantages, several limitations remain. First, the photometric stabilization stage operates offline and is not jointly optimized with reconstruction, suggesting future end-to-end formulations that jointly learn radiometric normalization and geometry. Pose estimation remains a bottleneck, as tools like COLMAP still underperform on stabilized thermal frames compared to RGB. Developing more robust, thermal-aware pose estimation methods is an important direction for future work.

Acknowledgements

This work was supported in part by the National Science Foundation under Grant No. IIS-2106786-001, and by the University of California, Riverside Regents' Faculty Fellowship. We thank Yi-Chun Hung for feedback on the manuscript and Mark Sheinin for insightful early conversations.

References

[1] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5855–5864, 2021. 2

[2] Abhay Singh Bhadoriya, Vamsi Vegamoor, and Sivakumar Rathinam. Vehicle detection and tracking using thermal cameras in adverse visibility conditions. *Sensors*, 22(12):4567, 2022. 1

[3] Spencer Carmichael, Manohar Bhat, Mani Ramanagopal, Austin Buchan, Ram Vasudevan, and Katherine A Skinner. Trnerf: Restoring blurry, rolling shutter, and noisy thermal images with neural radiance fields. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 7980–7990. IEEE, 2025. 3

[4] Qian Chen, Shihao Shu, and Xiangzhi Bai. Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis. In *European Conference on Computer Vision*, pages 253–269. Springer, 2024. 3, 6

[5] Zhilu Chen and Xinming Huang. Pedestrian detection for autonomous vehicle using multi-spectral cameras. *IEEE Transactions on Intelligent Vehicles*, 4(2):211–219, 2019. 1

[6] Mauricio Correa, Gabriel Hermosilla, Rodrigo Verschae, and Javier Ruiz-del Solar. Human detection and identification by robots using thermal and visual information in domestic environments. *Journal of Intelligent & Robotic Systems*, 66 (1):223–243, 2012. 1

[7] Hiba Dahmani, Moussab Bennehar, Nathan Piasco, Luis Roldao, and Dzmitry Tsishkou. Swag: Splatting in the wild images with appearance-conditioned gaussians. In *European Conference on Computer Vision*, pages 325–340. Springer, 2024. 2, 5

[8] Haixuan Ding, Jialiang Tang, Sheng Wan, and Chen Gong. Exploring neural radiance fields for thermal view synthesis solely with thermal inputs. *Chinese Journal of Electronics*, 2025. 3

[9] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5501–5510, 2022. 2

[10] Russell C Hardie and Douglas R Droege. A map estimator for simultaneous superresolution and detector nonunifomity correction. *J. Adv. in Signal Processing*, 2007:1–11, 2007. 2

[11] Russell C Hardie, Majeed M Hayat, Earnest Armstrong, and Brian Yasuda. Scene-based nonuniformity correction with video sequences and registration. *Appl. Optics*, 39(8):1241–1250, 2000. 2

[12] 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 preprint arXiv:2403.12154*, 2024. 1, 3, 6

[13] Lei He, Leheng Li, Wenchao Sun, Zeyu Han, Yichen Liu, Sifa Zheng, Jianqiang Wang, and Keqiang Li. Neural radiance field in autonomous driving: A survey. *arXiv preprint arXiv:2404.13816*, 2024. 1

[14] Zewei He, Yanpeng Cao, Yafei Dong, Jiangxin Yang, Yanlong Cao, and Christel-Löic Tisse. Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: A deep-learning approach. *Appl. Optics*, 57(18): D155–D164, 2018. 2

[15] Neham Jain, Andrew Jong, Sebastian Scherer, and Ioannis Gkioulekas. Smokeseer: 3d gaussian splatting for smoke removal and scene reconstruction. *arXiv preprint arXiv:2509.17329*, 2025. 3

[16] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering, 2023. 2, 6

[17] Jonas Kulhanek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. Wildgaussians: 3d gaussian splatting in the wild. *arXiv preprint arXiv:2407.08447*, 2024. 2, 5

[18] Ke Li, Mana Masuda, Susanne Schmidt, and Shohei Mori. Radiance fields in xr: A survey on how radiance fields are envisioned and addressed for xr research. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 1

[19] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5741–5751, 2021. 2

[20] Yvette Y Lin, Xin-Yi Pan, Sara Fridovich-Keil, and Gordon Wetzstein. Thermalnerf: Thermal radiance fields. In *2024 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2024. 1, 2, 3, 6

[21] Chengwei Liu, Xiubao Sui, Guohua Gu, and Qian Chen. Shutterless non-uniformity correction for the long-term stability of an uncooled long-wave infrared camera. *Measurement Science and Technology*, 29(2):025402, 2018. 2

[22] Rongfeng Lu, Hangyu Chen, Zunjie Zhu, Yuhang Qin, Ming Lu, Le Zhang, Chenggang Yan, and Anke Xue. Thermal-gaussian: Thermal 3d gaussian splatting. *arXiv preprint arXiv:2409.07200*, 2024. 1, 3

[23] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7210–7219, 2021. 2

[24] JR Martinez-De Dios and Anibal Ollero. Automatic detection of windows thermal heat losses in buildings using uavs. In *2006 world automation congress*, pages 1–6. IEEE, 2006. 1

[25] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf:

Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 6

[26] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16190–16199, 2022. 2

[27] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 2

[28] Robert Olbrycht and Bogusław Więcek. New approach to thermal drift correction in microbolometer thermal cameras. *Quantitative InfraRed Thermography Journal*, 12(2):184–195, 2015. 2

[29] Mert Özer, Maximilian Weiherer, Martin Hundhausen, and Bernhard Egger. Exploring multi-modal neural scene representations with applications on thermal imaging. In *European Conference on Computer Vision*, pages 82–98. Springer, 2024. 1, 2, 3, 6

[30] Manikandasriram Srinivasan Ramanagopal, Zixu Zhang, Ram Vasudevan, and Matthew Johnson-Roberson. Pixel-wise motion deblurring of thermal videos. *arXiv preprint arXiv:2006.04973*, 2020. 2

[31] Vishwanath Saragadam, Akshat Dave, Ashok Veeraraghavan, and Richard G Baraniuk. Thermal image processing via physics-inspired deep networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4057–4065, 2021. 2

[32] Ukcheol Shin, Kyunghyun Lee, Byeong-Uk Lee, and In So Kweon. Maximizing self-supervision from thermal image for effective self-supervised learning of depth and ego-motion. *IEEE Robotics and Automation Letters*, 7(3):7771–7778, 2022. 2

[33] R Vadivambal and Digvir S Jayas. Applications of thermal imaging in agriculture and food industry—a review. *Food and bioprocess technology*, 4(2):186–199, 2011. 1

[34] Guangming Wang, Lei Pan, Songyou Peng, Shaohui Liu, Chenfeng Xu, Yanzi Miao, Wei Zhan, Masayoshi Tomizuka, Marc Pollefeys, and Hesheng Wang. Nerfs in robotics: A survey. *The International Journal of Robotics Research*, page 02783649251374246, 2024. 1

[35] Xiangyu Wen, Guangchi Fang, Bo Yang, and Bing Wang. Geometry aware 3d multiview thermal reconstruction with emissive residual decomposition gaussian splatting. In *Proceedings of the 2025 ACM International Workshop on Thermal Sensing and Computing*, pages 7–12, 2025. 3

[36] Alejandro Wolf, Jorge E Pezoa, and Miguel Figueroa. Modeling and compensating temperature-dependent non-uniformity noise in IR microbolometer cameras. *Sensors*, 16(7):1121, 2016. 2

[37] Congrong Xu, Justin Kerr, and Angjoo Kanazawa. Splatfacto-w: A nerfstudio implementation of gaussian splatting for unconstrained photo collections. *arXiv preprint arXiv:2407.12306*, 2024. 2, 5

[38] Jiacong Xu, Mingqian Liao, Ram Prabhakar Kathirvel, and Vishal M Patel. Leveraging thermal modality to enhance reconstruction in low-light conditions. In *European Conference on Computer Vision*, pages 321–339. Springer, 2024. 1, 3

[39] Jiacong Xu, Yiqun Mei, and Vishal Patel. Wild-gs: Real-time novel view synthesis from unconstrained photo collections. *Advances in Neural Information Processing Systems*, 37:103334–103355, 2024. 2, 5

[40] Chongjie Ye, Yinyu Nie, Jiahao Chang, Yuantao Chen, Yihao Zhi, and Xiaoguang Han. Gaustudio: A modular framework for 3d gaussian splatting and beyond. *arXiv preprint arXiv:2403.19632*, 2024. 5

[41] Tianxiang Ye, Qi Wu, Junyuan Deng, Guoqing Liu, Liu Liu, Songpengcheng Xia, Liang Pang, Wenxian Yu, and Ling Pei. Thermal-nerf: Neural radiance fields from an infrared camera. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1046–1053. IEEE, 2024. 3, 5

[42] Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, et al. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025. 6

[43] Seokwon Yeom. Thermal image tracking for search and rescue missions with a drone. *Drones*, 8(2):53, 2024. 1

[44] Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. In *European Conference on Computer Vision*, pages 341–359. Springer, 2024. 2, 5

[45] Chonghao Zhong and Chao Xu. Tex-nerf: Neural radiance fields from pseudo-tex vision. *arXiv preprint arXiv:2410.04873*, 2024. 3

[46] Chen Zou, Qingsen Ma, Jia Wang, Rongfeng Lu, Ming Lu, and Zhaowei Qu. Tga-gs: Thermal geometrically accurate gaussian splatting. *Applied Sciences*, 15(9):4666, 2025. 1, 3