ThermoGS: Decoupling Physical Surface Attributes for Spatio-Temporal Thermal Field Emulation via 4D Gaussian Splatting

Kun Yang[1], Yuxiang Liu[2], Zeyu Cui[2], Shen Yan[2], Maojun Zhang[2], Yu Liu[2], Xue Wang[1(✉)], and Qing Wang[1]

[1] Northwestern Polytechnical University, Xi'an, China
xwang@nwpu.edu.cn
[2] National University of Defense Technology, Changsha, China

Abstract. Thermal infrared (TIR) imaging is indispensable for all-weather perception and energy management, yet existing 3D thermal modeling methods remain primarily confined to static environmental snapshots. These approaches typically struggle to capture complex temporal temperature variations and lack the capacity to simulate scenes under unseen environmental conditions. To bridge this gap, we propose ThermoGS, a unified framework that couples 4D Gaussian Splatting (4DGS) with fundamental thermodynamic principles to achieve physically-consistent thermal field emulation. Unlike previous data-driven methods that directly regress radiance, ThermoGS decouples intrinsic material properties, including emissivity (e), convective heat transfer coefficient (h), volumetric heat capacity (c), and absorptivity (α)—from transient temperature states. By integrating Physics-Informed Neural Networks (PINNs), our framework ensures that the predicted spatiotemporal thermal distribution strictly adheres to energy balance equations, enabling generalized simulation across arbitrary viewpoints and time periods. Furthermore, we introduce the Thermo Dataset, a comprehensive, drone-captured collection featuring synchronized multi-modal data and diverse environmental parameters. Extensive experiments demonstrate that ThermoGS significantly outperforms comparative methods in thermal simulation accuracy, maintaining a predicted temperature error within 1 degree Celsius. The code is available at https://github.com/NPU-CVPG/ThermoGS.

Keywords: Thermal Simulation · 4D Gaussian Splatting · Thermodynamics · Physics-Informed Neural Networks

1 Introduction

Thermal infrared (TIR) imaging offers a distinctive advantage for all-weather perception by enabling non-intrusive measurement of object surface temperatures, demonstrating robust adaptability to adverse environmental conditions

✉ Corresponding author.

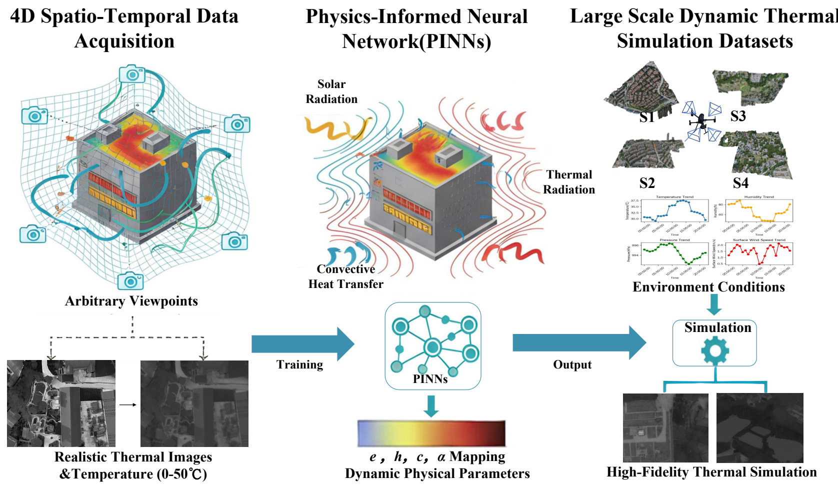

Fig. 1: Our framework integrates thermodynamic principles with 4D Gaussian Splatting for physically-consistent thermal emulation under clear-sky conditions. By inputting scene geometry and environmental factors (e.g., solar radiation, wind speed), it simulates 4D spatiotemporal thermal fields and renders accurate temperature maps from arbitrary views. Notably, ThermoGS decouples intrinsic material attributes to ensure adherence to thermodynamic laws.

such as low light and haze. These unique advantages have led to its widespread application in fields such as security surveillance, building inspection, and large-scale energy-efficiency analysis [8, 25, 29, 34]. With the rapid advancement of high-resolution, miniaturized TIR sensors and drone platforms, the foundation for modeling large-scale 3D thermal fields has been established. However, existing methodologies remain largely confined to reconstructing specific physical structures and static textures. We argue that the ultimate objective should transcend mere reconstruction; rather, it should aim for deep-learning-based 3D thermal scene simulation. The core challenge lies in generating physically accurate thermal distributions (e.g., temperature maps) under arbitrary, dynamic environmental conditions, a task that traditional reconstruction-centric frameworks fail to address.

Early research in thermal infrared processing [4–6, 10, 11, 15, 17, 22, 26, 28, 36] primarily focused on 3D scene reconstruction: they first built geometric models using RGB images and multi-view reconstruction techniques [9], then mapped pre-acquired TIR data onto these models to generate pseudo thermal scenes. While these methods facilitate visualization, they are fundamentally reconstruction-dependent. Specifically, they cannot independently simulate thermal distributions for unseen scenarios or fluctuating conditions, as they fail to leverage the intrinsic thermo-physical properties of objects to model dynamic temperature evolution.

Recent advancements in implicit representations, such as NeRF-based methods [7, 35], have extended thermal tasks to novel view rendering, but often suffer from geometric instability and limited adaptability to dynamic simulations. In contrast, 3D Gaussian Splatting (3DGS) [14] offers explicit scene modeling with superior scalability. While several GS-based thermal methods have emerged [2, 19], they remain limited to static snapshots and ignore temporal thermal variations driven by ecological factors such as wind speed, atmospheric pressure, and humidity. Notably, NTR-Gaussian [33] addressed dynamic thermal reconstruction at night but is limited to a single environmental condition and cannot predict full-day temperatures. Critically, these methods typically regress radiance directly without decoupling the underlying physical attributes, which limits their interpretability and generalization.

To address these limitations, we propose ThermoGS, a unified framework for generalized 3D thermal scene simulation that integrates fundamental thermodynamic principles with 4D Gaussian Splatting. In this work, we focus on modeling the dominant thermodynamic interactions under clear-sky conditions, where solar radiation serves as the primary energy driver. By taking scene geometry and localized environmental parameters as inputs, ThermoGS directly models a 4D spatiotemporal thermal field to generate realistic temperature maps across arbitrary viewpoints, time periods, and diverse environments (Fig. 1). Under the clear-sky assumption, we explicitly model energy balance involving solar radiation, convective heat transfer, and thermal radiation, leveraging Physics-Informed Neural Networks (PINNs) for physics-constrained learning. Unlike prior work, our model decouples intrinsic material properties-including emissivity (e), convective heat transfer coefficient (h), volumetric heat capacity (c), and absorptivity (α) from transient temperature states. This enables the thermal distribution to strictly follow thermodynamic laws. Additionally, we introduce a large-scale dynamic thermal simulation dataset that supports generalized simulation across diverse scenarios and environmental conditions.

Our contributions are summarized as follows:

- We propose ThermoGS, a unified simulation framework that bridges thermodynamic principles with 4DGS. By integrating PINNs, our method enables independent and generalized thermal simulations across diverse scenarios, environmental conditions, and time periods.
- We develop a physics-driven strategy that decouples intrinsic material properties at real-world scales. This ensures the physical authenticity of simulated thermal fields, which strictly adhere to thermodynamic laws under clear-sky conditions.
- We construct the Thermo Dataset, a comprehensive collection specifically designed for generalized 3D thermal scene simulation. The dataset contains approximately 40K infrared thermal images captured across diverse urban scenarios, synchronized with multi-dimensional environmental metadata such as wind speed, solar radiation, and atmospheric pressure.

2 Related Work

2.1 Thermal-Assisted 3D Reconstruction

Early efforts to integrate thermal information with 3D modeling focused on reconstructing geometrically accurate 3D scenes [3,12,23,24,27] with attached thermal attributes, rather than simulating thermal distributions. Rangel et al. [26] developed a multimodal system combining thermal cameras and depth sensors to fuse spatial and thermal data into a 3D thermal model, but this work relied on preacquired sensor data and lacked the ability to generate thermal information for unseen scenarios. Zhao et al. [36] improved the real-time performance of thermal reconstruction via multisensor calibration. In contrast, Muller et al. [22] proposed a T-ICP method that leverages the robustness of thermal data to viewpoint or illumination changes for the generation of high-fidelity models. Still, both methods were limited to static, data-dependent reconstruction and could not adapt to dynamic environmental conditions (e.g., varying temperatures and winds) or predict thermal dynamics over time. Li et al. [15] optimized stereo matching for thermal images to enhance reconstruction quality, but like other geometry-based approaches, it failed to model the intrinsic thermal-physical mechanisms (e.g., heat transfer, radiation) that govern real-world thermal behavior.

With the rise of implicit representation, NeRF-based [21] methods extended thermal reconstruction to novel view rendering. ThermoNeRF [7] and Thermal-NeRF [35] integrated thermal image ensembles with NeRF to generate 3D models with thermal information, but they inherited NeRF's limitations: slow rendering speed, unstable geometric representation, and reliance on large volumes of pre-collected TIR data. Most importantly, these methods remained reconstruction-oriented. They could only "reproduce" thermal patterns from existing data, not "simulate" thermal distributions for arbitrary scenarios, environmental conditions, or time periods, failing to meet the demands of generalized thermal simulation.

2.2 3D/4D Gaussian Splatting for Scene Modeling

3D Gaussian Splatting (3DGS) [14] emerged as a disruptive technique for 3D scene representation, addressing NeRF's inefficiencies via explicit, learnable Gaussian distributions. Its structured data format enables real-time novel-view rendering, scene editing, and easy integration with multimodal data (e.g., depth and LiDAR), making it a promising foundation for dynamic thermal tasks. Matsuki et al. [20], Keetha et al. [13], and Yan et al. [32] integrated depth cameras with 3DGS for SLAM, while Li et al. [16] and Zhu et al. [37] fused monocular depth estimates [1] with RGB data to refine geometric accuracy. These works validated the flexibility of 3DGS for multimodal fusion but focused solely on geometric reconstruction, without considering thermal factors.

4D Gaussian Splatting (4DGS) [31] extended 3DGS to dynamic scenes by introducing deformation fields, enabling reconstruction of time-varying geometries. However, it did not model thermal dynamics, as its core goal remained

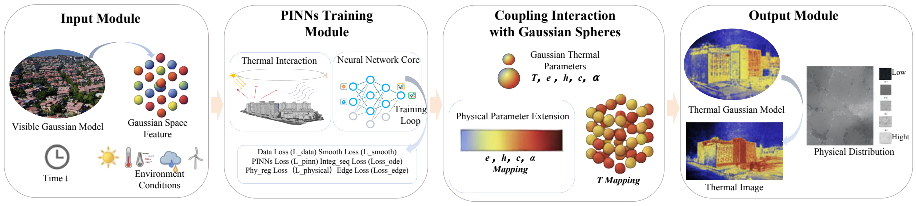

Fig. 2: Method Overview. The input of our model includes a Gaussian model reconstructed from visible-light images, Gaussian spatial features, current time, and environmental conditions (such as wind speed, temperature, humidity, solar radiation, and pressure). Employing 4D Gaussian splatting as the scene representation, we integrate PINNs and physical constraints of dynamic thermal interaction to approximate the partial differential equation governing the temperature field over time. Furthermore, we deeply couple the physical processes with the scale of Gaussian spheres, thereby simulating and obtaining accurate physical quantities, specifically e (emissivity), h (convective heat transfer coefficient), c (volumetric specific heat capacity), α (absorptivity), and the temperature T.

capturing spatiotemporal geometric changes rather than thermal-physical processes. Recent attempts to combine 3DGS with thermal data (e.g., [2, 18, 19]) achieved 3D thermal reconstruction, but they were limited to static spatial thermal mapping-they ignored environmental factors (e.g., ambient temperature, humidity) that drive thermal variation, did not predict thermodynamic parameters, and could not simulate thermal evolution over time. NTR-Gaussian [33] addressed nighttime dynamic reconstruction but suffers from poor generalization and is incapable of full-day temperature prediction

Our ThermoGS shifts the paradigm from reconstruction to generalized thermal simulation. By integrating 4DGS with thermodynamic principles and physics-constrained learning, it decouples intrinsic physical attributes (e.g., emissivity, heat capacity) to model thermal processes under clear-sky conditions, filling the gap between existing reconstruction methods and practical simulation needs.

3 Method

3.1 Overview

The pipeline of the proposed ThermoGS is shown in Fig. 2. Our method uses the reconstructed visible-light 3D Gaussian model, corresponding features, time, and environmental conditions (such as wind speed, temperature, humidity, solar radiation, and pressure). It utilizes a neural network to simulate the corresponding temperature values T and physical quantities e (emissivity), h (convective heat transfer coefficient), c (volumetric specific heat capacity), and α (absorptivity). During this process, we deeply couple Gaussian spheres with thermodynamics principles and PINNs to ensure the accuracy and authenticity of the predicted values.

3.2 Data Processing

Construction of the Visible Gaussian Model. First, we obtain the 3D visible light model of the drone flight area based on drone oblique photography technology, and manually label the ground object labels. Subsequently, we registered this model to infrared images to obtain accurate pose estimates. Then, we rendered visible images and constructed the visible 3D Gaussian model using the 3D visible model.
Acquisition of Gaussian Space Features. We convert visible 2D features into 3D Gaussian features with spatial information via Feature_GS [31], and assign a label to each Gaussian sphere.
Times and Environment Conditions. We determine the time based on the capture time of the thermal images. Then we normalize time to 0-1 using min-max normalization, then use a 10-order trigonometric function encoding to obtain the final encoded vector. We downloaded climate data corresponding to the target time from a global meteorological data website and used an interpolation algorithm to convert the discrete data to continuous values, thereby enabling the retrieval of environmental conditions at any arbitrary time.
Acquisition of Temperature Maps. To obtain accurate surface temperature T_{surf}, we apply radiometric correction to the raw thermal data. Specifically, we use the DJI Thermal SDK to compensate for atmospheric transmissivity and reflected background radiation based on synchronized meteorological data, rather than directly using the uncorrected apparent temperature.

3.3 Principle of Thermal Interaction

Under clear-sky conditions, the surface temperature T is governed by the immediate energy balance between the object and its environment. Since we employ radiometric correction via the DJI Thermal SDK to compensate for atmospheric transmissivity and reflected background radiation during data preprocessing, the "ground truth" labels effectively represent the intrinsic surface temperature. Consequently, we focus on the dominant first-order thermodynamic interactions. The simplified theory of thermal interaction is based on thermodynamic principles, as shown in Fig. 3. During the daytime, the scene's thermal temperature is determined by both absorption and release processes, with absorption mainly from solar radiation and release mainly from convective heat exchange and radiation. At night, the outdoor scene's thermal temperature is mainly determined by the release process. We define the scene surface temperature T as a model related to time t, environmental conditions C, the model's relative position P, and model features F, expressed in Eq. 1,

$$T = f(t, C, P, F). \tag{1}$$

The convective heat exchange and radiative heat release processes are represented by Eq. 2 and Eq. 3, respectively. Both of these processes occur continuously.

$$Q_{conv} = hA_{\mathrm{surf}}(T - T_m), \tag{2}$$

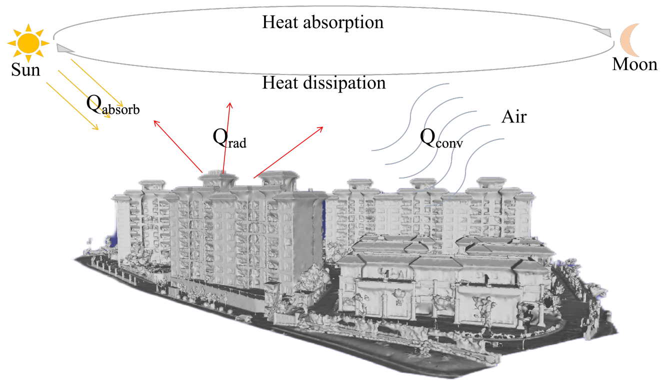

Fig. 3: Thermal Interaction Theory. The temperature of outdoor scenes exhibits heat absorption and heat release characteristics throughout the day and night. During the entire day, the scene releases heat mainly through its own radiation and convective heat transfer; while in the daytime, objects additionally absorb heat from solar radiation. Theoretical interaction model is inspired by [33].

where h represents the convective heat transfer coefficient of different objects in the scene, A_{surf} represents the heat exchange surface area, T represents the radiative temperature at each moment, and T_m represents the environmental temperature at each moment. This formula calculates the amount of heat exchanged between the scene and the air through convection within a unit of time.

$$Q_{rad} = e\sigma A_{\text{rad}} T^4. \tag{3}$$

Here, e represents the emissivity of different objects in the scene, σ is the Stefan-Boltzmann constant, A_{rad} represents the thermal radiation surface area, and T represents the radiative temperature at each moment. This formula calculates the amount of heat radiated by the scene within a unit of time.

During the daytime, in addition to the above processes, there is also a process of absorbing heat from solar radiation, which can be expressed as Eq. 4,

$$Q_{absorb} = \alpha A_{\perp} G_{sun}. \tag{4}$$

Among them, α denotes the absorptivity, $A_{\perp}$denotes the projected area of the scene in the direction of solar incidence, and G_{sun} denotes the solar radiation intensity. This formula represents the heat absorbed by the scene per unit time.

By integrating the processes above, we can derive the PDE of temperature T with respect to time t in Eq. 5.

$$V * c * \frac{\partial T}{\partial t} = Q_{absorb} - Q_{rad} - Q_{conv}, \tag{5}$$

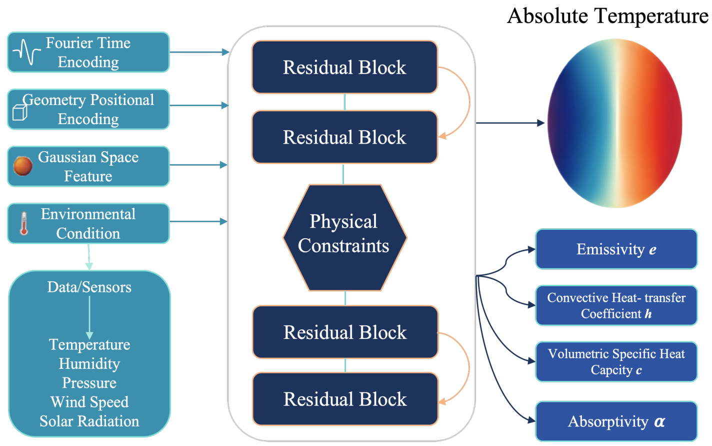

Fig. 4: Core Architecture. The model inputs include Fourier positional encoding of time, point geometric positional encoding, spatial features, and external environmental quantities (ambient temperature, relative humidity, air pressure, wind speed, and solar irradiance), with temperature and physical quantities as outputs.

where c represents the volumetric specific heat capacity and V represents the scene volume. Solving the equation will enable us to calculate the object's radiative temperature at any given moment.

3.4 Thermal Interaction-Gaussian Sphere Coupling Training

Motivation for Continuous Physical Field Learning. While categorical lookup tables offer a simplified baseline for parameter assignment, they fail to account for the spatial heterogeneity inherent in real-world urban materials. For instance, a single "building" label encompasses diverse surfaces with varying degrees of aging, roughness, and structural orientations, all of which significantly influence local thermodynamic response. Instead of relying on rigid, discrete values, our ThermoGS leverages PINNs to infer a continuous physical manifold. By predicting physical quantities (e,h,c,α) at a per-Gaussian level, the framework can adaptively refine material properties based on the observed spatiotemporal temperature evolution. This approach ensures that the simulated thermal field is not merely a piece-wise constant approximation but a high-fidelity representation of the scene's intrinsic thermodynamic complexity.
Thermal Interaction Integrating Gaussian Sphere Scale. In the process of thermal interaction, the terms associated with Gaussian spheres in Eq. 2, 3, 4, and 5 are A_{surf}, A_{rad}, $A_{\perp}$, and V respectively. These can thus be integrated into Gaussian ellipsoids along with depth, ensuring that the thermal interaction is fully consistent with the real scale. After simplifying the model, the calculations of A_{surf}, A_{rad}, $A_{\perp}$ and V can be expressed as,

$$A_{\text{surf}} = \frac{\pi}{3}(ab + ac + bc), \tag{6}$$

$$A_{\text{rad}} = 4\pi \left(\frac{a^p b^p + a^p c^p + b^p c^p}{3} \right)^{1/p}, \tag{7}$$

$$A_{\perp} = \pi \frac{abc}{((ad_x)^2 + (bd_y)^2 + (cd_z)^2)^{\frac{1}{2}}}, \tag{8}$$

$$V = \frac{4}{3}\pi abc. \tag{9}$$

Among them, a, b, and c represent the semi-major axes of the Gaussian ellipsoid; d_x, d_y, and d_z represent the unit vector of the solar incident direction in the local coordinate system of the ellipsoid; and p is a constant, approximately 1.61.

Training Process. As shown in Fig. 4, we adopt a unified physics-aware network to predict the absolute temperature of Gaussian ellipsoid instances directly. The input consists of Fourier positional encoding of time, point geometric positional encoding, spatial features, and external environmental quantities (ambient temperature, air pressure, wind speed, relative humidity, and solar irradiance); all environmental quantities are derived from sensors and used as network inputs rather than network outputs. The network is a deepened multi-layer residual MLP, with several parallel physical property heads to predict the convective heat transfer coefficient, emissivity, volumetric heat capacity/specific heat, absorptivity, and latent heat coefficient, which are used to construct physical constraints. Energy balance is computed at the absolute scale during the network's forward pass.

3.5 Loss Function

Supervision is derived from the reconstruction loss L_{data} (including L_1 and SSIM loss) and edge loss L_{edge} of multi-view infrared temperature maps at different time steps. Physical induction is incorporated via auxiliary losses:

Finite Difference PINN Loss L_{pinn}: Aligns the numerical difference-based temperature change between adjacent time pairs with the consistency of the network-derived $\frac{\partial T}{\partial t}$.

Physical Consistency Loss $L_{physical}$: Looks up the values of various physical quantities from a table, and uses real physical quantities for regularization to constrain the model's physical processes, ensuring that the predicted physical quantities are close to the real values.

ODE Integration Consistency Loss L_{ode}: Aligns the temperature trajectory derived from $\frac{\partial T}{\partial t}$ with the directly predicted temperature, and re-evaluates physical properties at the midpoint.

Temporal Smooth Loss L_{smooth}: A weak regularization term that prevents abrupt temperature changes between adjacent time periods. It is defined as:

$$L_{smooth} = \sum_{j=1}^{N-1} \left(\frac{1}{M} \sum_{i=1}^{M} |\hat{T}_{j+1,i} - \hat{T}_{j,i}| \right), \tag{10}$$

Table 1: Detailed Statistics of the Thermo Dataset. Our dataset covers 10 land-cover categories: Building, Road, Vehicle, Grassland, High Vegetation, Sports Field, Water Body, Farmland, Construction Site, and Bare Ground.

Scene	Main Categories	Time	Frames	Split
S1 (Substation)	Build., Cons., Bare.	Full-day	10K	9/1
S2 (School)	Build., Grass., Sport.	Full-day	10K	
S3 (Square)	Road, Water, Veg.	Full-day	10K	
S4 (Driving school)	Road, Vehicle, Bare.	Full-day	10K	
Total	10 Categories	Full-day	40K	9/1

where M represents the number of Gaussian spheres, and N represents the number of differential intervals.

The overall objective is to minimize the weighted sum of image reconstruction losses and physical constraint losses, enabling the network to maintain temperature evolution consistent with real physical dimensions and diurnal processes even when supervised solely by infrared images. It is given as follows,

$$L = \lambda_1 L_{data} + \lambda_2 L_{edge} + \lambda_3 L_{smooth} + \lambda_4 L_{physical} + \lambda_5 L_{ode} + \lambda_6 L_{pinn}, \quad (11)$$

where we set $\lambda_1 = 0.8$, $\lambda_2 = 0.2$, $\lambda_3 = 1.0$, $\lambda_4 = 0.01$, $\lambda_5 = 0.01$, $\lambda_6 = 1.0$.

4 Experiments

4.1 Experiment Setup

Thermo Datasets. To evaluate methods for dynamic thermal simulation, we have developed the Thermo dataset as shown in Fig. 1. As illustrated in Tab. 1, This dataset covers four distinct scenes. For each scene, we used the DJI Dock Matrice 4TD for all-weather TIR image capture, and then selected several representative days for method validation. The flight altitude is set at 150 meters. The thermal imaging sensor ha a focal length of 12.0 mm, and a temperature range from -40°C to 550°C. These images effectively capture the temporal variations in thermal radiation from objects, such as buildings, under nighttime conditions, allowing the TIR images with timestamps to serve as representative samples of dynamic temperature changes in the scene. For environmental conditions, we select the four factors with the greatest influence on temperature changes as inputs: wind speed, temperature, humidity, solar radiation, and pressure. The dataset comprises: (1) **Images**: 640×512 TIR frames with a GSD of $\sim$15 cm/pixel. (2) **Visible**: High-resolution RGB images (1920×1080) for initial 3D reconstruction. (3) **Meteorological**: Environmental metadata synchronized with each TIR frame.

Data Division. Our Thermo dataset contains infrared images captured across different scenes, months, and times of day. We used specific time periods from some scenes as the training set, while other time periods from additional scenes

Table 2: Quantitative Comparison of Thermal Field Simulation Accuracy. We evaluate our ThermoGS against state-of-the-art methods under two rigorous protocols: **Cross-Time Generalization**, where models are tested on unseen time periods within the same scene, and **Cross-Scene Generalization**, where models are tested on entirely unseen scenarios. Metrics include PSNR (dB), SSIM, and Temperature MAE ($^{\circ}C$). Bold indicates the best performance.

Method	S1 (S2,S3,S4)			S2 (S1,S3,S4)			S3 (S1,S2,S4)			S4 (S1,S2,S3)			S1-S4		
	PSNR↑	SSIM↑	MAE↓	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓
4DGS [31]	24.34	0.866	2.147	29.88	0.941	1.067	23.66	0.949	2.533	27.37	0.934	1.403	28.82	0.948	1.475
ThermalGS [18]	24.38	0.859	2.199	30.48	0.948	1.022	24.22	0.952	2.333	27.42	0.944	1.426	29.69	0.948	1.396
NTR-Gaussian [33]	25.03	0.872	1.993	27.39	0.940	1.524	21.78	0.944	3.122	28.80	0.950	1.278	31.42	0.956	1.136
Ours	**26.82**	**0.874**	**1.464**	**31.24**	**0.948**	**0.890**	**26.03**	**0.955**	**1.49**	**28.81**	**0.954**	**1.173**	**31.83**	**0.960**	**0.932**

Table 3: Quantitative comparison on the NTR Dataset. We use PSNR, SSIM, and MAE as the quantitative indicators for the four regions respectively, and calculate the average value of the performance indicators.

Method	S1			S2			S3			S4		
	PSNR↑	SSIM↑	MAE↓	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓	PSNR ↑	SSIM ↑	MAE ↓
4DGS [31]	32.19	0.961	1.655	29.17	0.961	2.475	29.78	0.954	2.224	32.06	0.958	1.616
Thermal3D-GS [2]	34.03	0.957	1.106	30.61	0.960	1.840	33.26	0.968	1.276	31.79	0.956	1.432
NTR-Gaussian [33]	34.50	0.962	0.947	31.43	0.964	1.709	35.15	0.972	0.959	34.61	0.964	1.065
Ours	**34.62**	**0.975**	**0.824**	**31.85**	**0.972**	**0.785**	**35.41**	**0.983**	**0.752**	**34.94**	**0.978**	**0.896**

served as the test set to ensure the generalization performance and accuracy of our method. Furthermore, we normalized the infrared thermal data into temperature maps.

Evaluation Metrics. Following the evaluation metrics in 3DGS [14] and temperature evaluation metrics, we use PSNR, SSIM [30], and Mean Absolute Error (MAE) of temperature values as the evaluation metrics.

Implementation Details. During the training process, we downsampled the original point cloud of each scene to ensure spatial efficiency. Our experimental environment consists of two NVIDIA RTX 4090 GPUs (24 GB memory), and the resolution of both training and testing images was set to 640×512. Additional details, including the dataset used, training configurations, and other information, are available in our supplementary materials. The GPU memory and inference speed are shown in Tab. 9, and our method is within a reasonable range.

4.2 Comparison Experiment

Comparison Methods. We divided our evaluation into three strategies. First, we trained our model on all scenario training sets to verify the generalization performance across different times and environmental conditions. Second, we conducted cross-scenario evaluation by excluding one scene during training and testing on the remaining one to assess spatial transferability. Third, to specifically verify the performance under purely nocturnal conditions, we evaluated

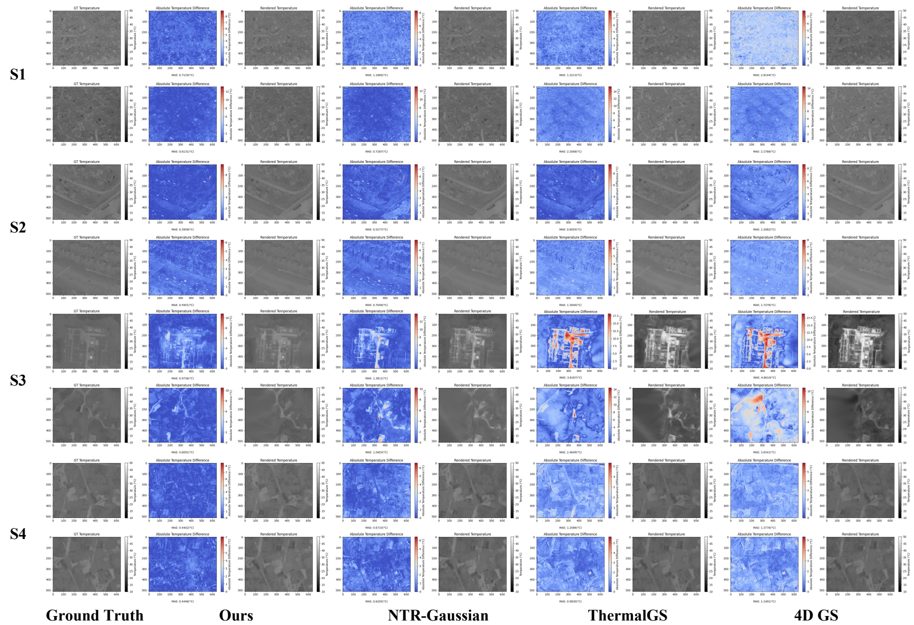

Fig. 5: Qualitative comparison. The leftmost column displays the ground truth (GT) temperature images. On the right side of the ground truth, the right column for each method shows the predicted temperature images. In contrast, the left column visualizes the error between the ground truth and predicted temperatures. The error has been normalized using min-max normalization and mapped to a red-to-blue color scale, where regions closer to red indicate larger absolute temperature errors, and regions closer to blue indicate smaller errors. Our method outperforms others in the distribution of temperature differences across the entire scene and in Mean Absolute Error (MAE).

ThermoGS on the original nighttime sequences provided by NTR-Gaussian [33]. To ensure fairness, all methods were trained and tested with identical settings on the same dataset.

Results. As shown in Tab. 2 and Tab. 3, our ThermoGS outperforms other methods in quantitative metrics, with the predicted temperature error within 1 degree Celsius. Figure 5 presents a visual comparison of the MAE of temperature. Our method achieves more accurate temperature predictions. It is important to note that NTR-Gaussian [33] was explicitly designed for nighttime reconstruction and does not model solar radiation. Including it in our full-day evaluation (Tab. 2) is not to highlight a failure of their method, but rather to empirically demonstrate the absolute necessity of modeling solar absorption for full-day spatio-temporal simulation. For a strict and fair comparison, Tab. 3 evaluates both methods exclusively on nighttime sequences, where our framework still achieves superior temperature prediction accuracy.

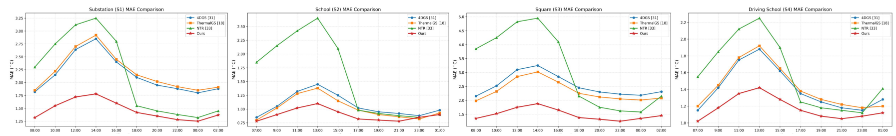

Fig. 6: Temporal Consistency Evaluation. Our ThermoGS achieves lower temperature errors across all time slots by leveraging thermodynamic constraints.

Table 4: Training and Inference Overhead. All measurements are collected on a single RTX 4090 GPU, including training wall-clock time, VRAM consumption and per-sample inference latency.

Method	Training (h)	VRAM (GB)	Infer Time (s)
4DGS [31]	1	15	6
NTR-Gaussian [33]	9	33	13
Ours	24	20	12

Temporal Consistency. We evaluate temporal consistency via temperature MAE across 10 consecutive time intervals with overlapping viewpoints. As shown in Fig. 6, our method achieves consistently lower temperature errors under both daytime and nighttime conditions, verifying that the incorporated thermodynamic constraints enable ThermoGS to produce high-fidelity and temporally stable thermal sequences. Moreover, across all evaluated scenarios, the thermal prediction MAE consistently peaks at 13:00–14:00. This trend aligns with the physical diurnal heating pattern: peak solar irradiance and ambient temperature in the early afternoon induce complex thermodynamic responses on road and building surfaces. Such intense thermal variations bring greater prediction challenges for generative models, leading to the maximum MAE at noon, while the error gradually decreases as the environment cools down.

Training Overhead. Training and inference take longer because ThermoGS must solve thermodynamic PDEs (Eq. 5) to ensure physical realism. While 4DGS [31] is faster, it only performs pixel interpolation and lacks physical constraints. ThermoGS enables physically-consistent emulation and simulation across diverse climates by decoupling intrinsic properties. For thermal simulation, physical integrity is more critical than raw inference speed.

4.3 Ablation Study

We ablated each component of the loss function to demonstrate the effectiveness of every part in our method, including L_{smooth}, $L_{physical}$, L_{pinn}, L_{ode} and L_{edge}. To illustrate, we present the quantitative results of the first strategy of the comparative experiments in Tab. 5. Our ThermoGS achieves optimal performance when all components are included and outperforms other strategies on quantitative metrics. As shown in Fig 7, the temperature error of our model is the smallest when all loss terms are present.

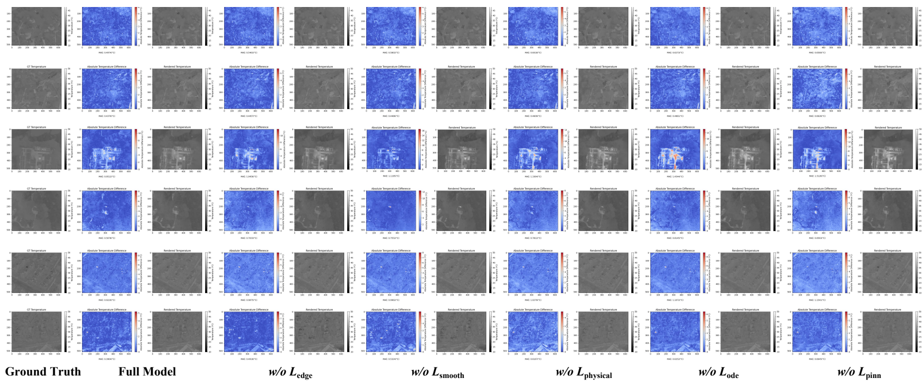

Fig. 7: We use image mean absolute error (MAE) to measure the accuracy of temperature prediction. Visual comparisons are conducted under the scenarios of the complete model and models without edge loss, smoothness loss, physical supervision loss, ODE loss, and PINN loss, respectively. Experimental results show that our strategy is optimal.

Table 5: Quantitative ablation study. We report PSNR, SSIM, and MAE for evaluating the rendering quality

L_{edge}	L_{smooth}	$L_{physical}$	L_{ode}	L_{pinn}	PSNR↑	SSIM↑	MAE↓
	✓	✓	✓	✓	31.17	0.957	1.147
✓		✓	✓	✓	31.27	0.957	1.163
✓	✓		✓	✓	31.33	0.956	1.147
✓	✓	✓		✓	31.42	0.956	1.175
✓	✓	✓	✓		31.42	0.956	1.133
✓	✓	✓	✓	✓	**31.83**	**0.960**	**0.932**

4.4 Physical Quantity Prediction

Since Thermo uses physical loss to constrain the prediction of physical quantities and combines physical formulas with the true scale of Gaussian spheres, our predictions are close to the physically actual values. As shown in Tab. 6, temperature calculation based on predicted physical quantities yields a smaller error compared with direct table lookup.

We obtained reference values from the table and derived approximate physical quantities for the different land cover types. Detailed values are shown in Tab. 8. We calculated the MAE between the predicted physical quantities and the actual reference values, and compared the results with those of NTR-Gaussian [33]. As shown in Tab. 7, our framework effectively learns to balance these semantic priors with transient temperature observations. While the generic lookup values represent class-level averages and do not capture intra-class variance, minimizing the distance to these soft targets via $L_{physical}$ ensures the optimization remains within physically plausible manifolds.

Table 6: Physical Quantity Ablation Study. The importance of predicting physical quantities.

Table Lookup	Inferred Attributes	PSNR↑	SSIM↑	MAE↓
✓		31.06	0.954	1.282
	✓	**31.83**	**0.960**	**0.932**

Table 7: Physical Attribute Inference. MAE between predicted and reference values (Tab. 8) across 10 categories.

Method	$\mathrm{MAE}_\epsilon \downarrow$	$\mathrm{MAE}_h \downarrow$	$\mathrm{MAE}_c \downarrow$	$\mathrm{MAE}_\alpha \downarrow$
NTR-Gaussian [33]	0.470	32.80	160.80	-
Ours	**0.018**	**0.534**	**5.86**	**0.019**

Table 8: Based on the manually annotated land cover types, we look up the table to calculate the approximate values of physical quantities for different land covers.

Category	e	α	h (W/(m²·K))	c (kJ/(m³·K))
Building	0.85	0.65	7	2400
Road	0.88	0.88	8	2100
Vehicle	0.85	0.70	15	3500
Grassland	0.92	0.75	7	2800
High Vegetation	0.93	0.78	8	3000
Sports Field	0.85	0.80	9	2200
Water Body	0.96	0.65	15	4200
Farmland	0.88	0.75	7	2000
Construction Site	0.82	0.65	8	1800
Bare Ground	0.85	0.70	6	1900

Table 9: We detail the VRAM consumption and inference speed compared to the baselines, showing our approach accomplishes its goals with acceptable computational overhead.

Method	GPU Memory ↓	inference FPS ↑
4DGS [31]	15GB	6
Thermal3D-GS [2]	7GB	14
NTR-Gaussian [33]	33GB	13
ThermalGS [18]	18GB	7
Ours	20GB	6

5 Conclusion

By integrating thermodynamic constraints with the absolute geometric scale of Gaussian ellipsoids, ThermoGS facilitates the inference of a continuous physical manifold rather than a piecewise-constant field. This allows our framework to adaptively refine parameters to account for local spatial heterogeneity-such as material aging and surface roughness which is typically ignored by categorical lookup tables.

To evaluate the accuracy of the inferred attributes, we obtained empirical reference values for 10 distinct land-cover categories. We then calculated the Mean Absolute Error (MAE) between our per-Gaussian predictions and these empirical priors. As shown in Tab. 8, ThermoGS significantly outperforms NTR-Gaussian [33] across all thermodynamic dimensions. Specifically, our method achieves a substantial reduction in MAE for emissivity, convective heat transfer coefficient, volumetric heat capacity, and absorptivity. These results demonstrate that our physics-informed learning strategy effectively captures the intrinsic thermodynamic properties of the scene, ensuring that the simulated temperature field is both visually realistic and physically grounded.

Acknowledgements. This work was supported by NSFC under Grant 62571449.

References

1. Bhat, S.F., Birkl, R., Wofk, D., Wonka, P., Müller, M.: Zoedepth: Zero-shot transfer by combining relative and metric depth. arXiv preprint arXiv:2302.12288 (2023)
2. Chen, Q., Shu, S., Bai, X.: Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis. arXiv preprint arXiv:2409.08042 (2024)
3. Dai, A., Nießner, M., Zollhöfer, M., Izadi, S., Theobalt, C.: Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. ACM Transactions on Graphics (ToG) **36**(4), 1 (2017)
4. Dlesk, A., Vach, K., Pavelka, K.: Transformations in the photogrammetric co-processing of thermal infrared images and rgb images. Sensors **21**(15), 5061 (2021)
5. Dlesk, A., Vach, K., Pavelka, K.: Photogrammetric co-processing of thermal infrared images and rgb images. Sensors **22**(4), 1655 (2022)
6. Ham, Y., Golparvar-Fard, M.: An automated vision-based method for rapid 3d energy performance modeling of existing buildings using thermal and digital imagery. Adv. Eng. Inform. **27**(3), 395–409 (2013)
7. Hassan, M., Forest, F., Fink, O., Mielle, M.: Thermonerf: Multimodal neural radiance fields for thermal novel view synthesis. arXiv preprint arXiv:2403.12154 (2024)
8. Hoegner, L., Stilla, U.: Mobile thermal mapping for matching of infrared images with 3d building models and 3d point clouds. Quant. Infrared Thermogr. J. **15**(2), 252–270 (2018)
9. Jadhav, T., Singh, K., Abhyankar, A.: A review and comparison of multi-view 3d reconstruction methods. Journal of Engineering Research **5**(3) (2017)
10. Javadnejad, F.: Small unmanned aircraft systems (UAS) for engineering inspections and geospatial mapping. Ph.D. thesis, Oregon State University (2017)
11. Jia, T., Wang, H., Chen, D., Hou, M., Wu, C., Tu, M., Jiang, Y., Zhang, S.: 3d temperature distribution model based on thermal infrared image. Journal of Sensors **2017**(1), 4815021 (2017)
12. Kähler, O., Prisacariu, V.A., Ren, C.Y., Sun, X., Torr, P., Murray, D.: Very high frame rate volumetric integration of depth images on mobile devices. IEEE Transactions on Visualization and Computer Graphics **21**(11), 1241–1250 (2015)
13. Keetha, N., Karhade, J., Jatavallabhula, K.M., Yang, G., Scherer, S., Ramanan, D., Luiten, J.: Splatam: Splat track & map 3d gaussians for dense rgb-d slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21357–21366 (2024)
14. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)
15. Li, H., Wang, S., Bai, Z., Wang, H., Li, S., Wen, S.: Research on 3d reconstruction of binocular vision based on thermal infrared. Sensors **23**(17), 7372 (2023)
16. Li, J., Zhang, J., Bai, X., Zheng, J., Ning, X., Zhou, J., Gu, L.: Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20775–20785 (2024)
17. Lin, D., Bannehr, L., Ulrich, C., Maas, H.G.: Evaluating thermal attribute mapping strategies for oblique airborne photogrammetric system aos-tx8. Remote Sens. **12**(1), 112 (2019)
18. Liu, Y., Chen, X., Yan, S., Cui, Z., Xiao, H., Liu, Y., Zhang, M.: Thermalgs: Dynamic 3d thermal reconstruction with gaussian splatting. Remote Sensing **17**(2), 335 (2025)

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. Matsuki, H., Murai, R., Kelly, P.H., Davison, A.J.: Gaussian splatting slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18039–18048 (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, A.R.O.: Close range 3D thermography: real-time reconstruction of high fidelity 3D thermograms, vol. 7. Kassel University Press GmbH (2019)
23. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: 2011 10th IEEE international symposium on mixed and augmented reality. pp. 127–136. Ieee (2011)
24. Nießner, M., Zollhöfer, M., Izadi, S., Stamminger, M.: Real-time 3d reconstruction at scale using voxel hashing. ACM Transactions on Graphics (ToG) **32**(6), 1–11 (2013)
25. Ramón, A., Adán, A., Castilla, F.J.: Thermal point clouds of buildings: A review. Energy Build. **274**, 112425 (2022)
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. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4104–4113 (2016)
28. Vidas, S., Moghadam, P.: Heatwave: A handheld 3d thermography system for energy auditing. Energy Build. **66**, 445–460 (2013)
29. Vidas, S., Moghadam, P., Sridharan, S.: Real-time mobile 3d temperature mapping. IEEE Sens. J. **15**(2), 1145–1152 (2014)
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. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4d gaussian splatting for real-time dynamic scene rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20310–20320 (2024)
32. Yan, C., Qu, D., Xu, D., Zhao, B., Wang, Z., Wang, D., Li, X.: Gs-slam: Dense visual slam with 3d gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19595–19604 (2024)
33. Yang, K., Liu, Y., Cui, Z., Liu, Y., Zhang, M., Yan, S., Wang, Q.: Ntr-gaussian: Nighttime dynamic thermal reconstruction with 4d gaussian splatting based on thermodynamics. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 691–700 (2025)
34. Yang, M.D., Su, T.C., Lin, H.Y.: Fusion of infrared thermal image and visible image for 3d thermal model reconstruction using smartphone sensors. Sensors **18**(7), 2003 (2018)
35. 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. arXiv preprint arXiv:2403.10340 (2024)

36. 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)
37. Zhu, Z., Fan, Z., Jiang, Y., Wang, Z.: Fsgs: Real-time few-shot view synthesis using gaussian splatting. In: European Conference on Computer Vision. pp. 145–163. Springer (2025)