DTGS: Physics-Embedded Dynamic Thermal 3D Reconstruction with Gaussian Splatting

Junbo Li[1], **Hang Fu**[1], **Yinuo Wang**[1], **Weimin Yuan**[2],
Cai Meng[1,3]*, **Xiangzhi Bai**[1,3,4]†
[1]School of Astronautics, Beihang University, Beijing, China
[2]Faculty of Intelligence Technology, Shanghai Institute of Technology, Shanghai, China
[3]Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technology, Ministry of Education, Beijing, China
[4]State Key Laboratory of Virtual Reality Technology and Systems
Beihang University, Beijing China

Abstract

Dynamic scene representation and rendering in the visible spectrum has been extensively studied. Compared to visible-light imaging, thermal infrared sensing offers all-weather observation and strong penetration capability, enabling perception in low-visibility environments. However, dynamic reconstruction in thermal infrared scenes is challenged due to atmospheric transport effects and interference from low-emissivity background regions, which often leads to motion blur and details loss in rendered results. To address these issues, we propose DTGS, a physics-driven dynamic 3D scene reconstruction approach. DTGS factorizes dynamic thermal scenes into radiometric and geometric components, modeling radiometric changes via rendering that integrates atmospheric radiative transfer, while capturing geometric deformations through Adaptive Temporal Gaussian basis. In addition, we design a thermal-radiance-weighted structural similarity loss to suppress gradient interference from low-radiance noisy regions. Furthermore, to demonstrate the effectiveness of our method, a large-scale dataset for this field named Dynamic_LTR is created. Experimental results demonstrate that our method achieves a 2.7 dB improvement in PSNR, a twofold speedup in rendering, and a threefold reduction in training time. The code and datasets will be available at `https://github.com/Bo3D777/DTGS`.

1 Introduction

Dynamic 3D scene reconstruction from video sequences serves as a critical foundation task for robotic navigation [1, 2, 3], autonomous driving [4, 5], and security surveillance [6, 7]. In recent years, dynamic reconstruction and novel view synthesis based on NeRF [8] or 3DGS [9] in visible-light scenarios have advanced rapidly [10, 11, 12, 13]. However, RGB imaging heavily depends on environmental illumination and weather conditions, leading to significant degradation in observation quality under low-visibility scenarios. In contrast, thermal infrared imaging (TIR imaging) primarily relies on the thermal radiation emitted by objects themselves, offering stronger all-weather and low-light perception capabilities, which can provide stable observations when visible light is unreliable.

But the inherent properties of thermal infrared imaging pose significant challenges for high-quality dynamic 3D reconstruction (see Fig. 1). In dynamic scenes, atmospheric transmission and the

*Primary corresponding author. Email: tsai@buaa.edu.cn
†Secondary corresponding author. Email: jackybxz@buaa.edu.cn

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

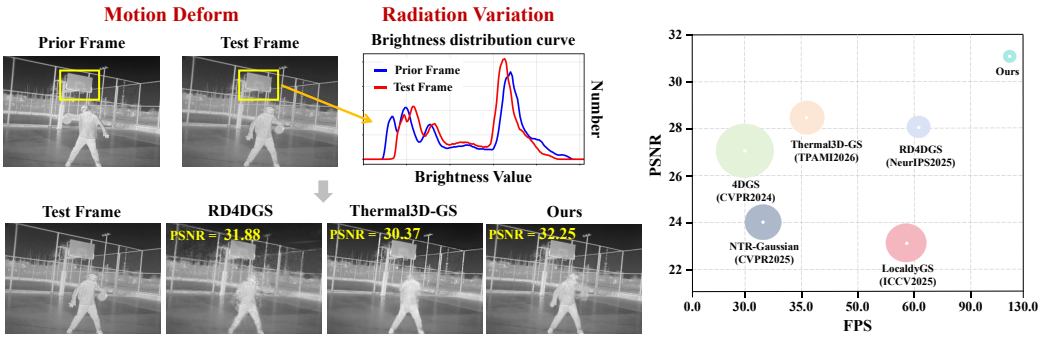


Figure 1: Our framework achieves significant advantages over state-of-the-art methods in terms of accuracy, rendering speed, and training time for thermal infrared dynamic scene reconstruction. In the table on the right, the larger circle area indicates the longer required training time.

radiation intensity of endogenous heat sources (e.g. pedestrians, vehicles) varies over time, causing inconsistencies in object features across frames and resulting in feature ambiguity in reconstruction. Moreover, low signal-to-noise ratios in background regions introduce noise that can hinder the accurate reconstruction of foreground dynamic objects. To address these challenges, this paper proposes a physics-embedded dynamic thermal 3D reconstruction approach, termed DTGS (see Fig. 2). DTGS fully exploits the physical characteristics of thermal infrared imagery by integrating the radiative transfer equation with the rasterization rendering formulation, enabling accurate modeling of radiation intensity variations within the scene. In addition, we introduce an Adaptive Temporal Gaussian basis to efficiently and precisely model geometric deformations. Furthermore, we design a thermal-radiance-weighted structural similarity loss that suppresses noise in low-radiance regions while enhancing reconstruction fidelity of dynamic, heat-emitting foreground objects.

To validate the proposed approach, we construct a dynamic thermal infrared 3D reconstruction dataset named Dynamic_LTR, comprising 31 indoor and outdoor scenes. The outdoor subset includes challenging low-visibility conditions—such as nighttime, overcast weather, and heavy fog—where RGB imaging is unreliable. To the best of our knowledge, Dynamic_LTR is currently the largest and most diverse dataset for dynamic thermal infrared 3D reconstruction, featuring complex scene variations and motions. Extensive experiments on Dynamic_LTR demonstrate the effectiveness of DTGS. Across all 31 scenes, DTGS achieves a 2.7 dB improvement in average PSNR, a twofold speedup in rendering, and a threefold reduction in training time.

Our main contribution can be summarized as follows:

(1) A physics-embedded dynamic thermal infrared 3D reconstruction method, termed DTGS, is proposed. It effectively integrates the physical properties of thermal infrared imaging with the 3DGS framework, thereby addressing challenges arising from atmospheric transmission effects and background noise.

(2) A large-scale dataset for dynamic thermal infrared 3D reconstruction termed Dynamic_LTR is proposed. To the best of our knowledge, it is currently the largest and most diverse of its kind, encompassing 31 scenes across 14 distinct categories, and can serve as a comprehensive benchmark for this field.

(3) Extensive experiments are conducted on the Dynamic_LTR dataset to evaluate the performance of current mainstream methods. The results demonstrate that DTGS achieves a 2.7 dB improvement in average PSNR, a twofold speedup in rendering, and a threefold reduction in training time.

2 Related Work

2.1 Dynamic 3D Scene Reconstruction

Accurately reconstructing the motion trajectories of dynamic scenes is an important and challenging task. In previous years, many dynamic reconstruction algorithms based on NeRF have emerged, such as deformation fields [14, 15, 16], scene flow [17, 18], and multi-plane representations [19, 20, 11].

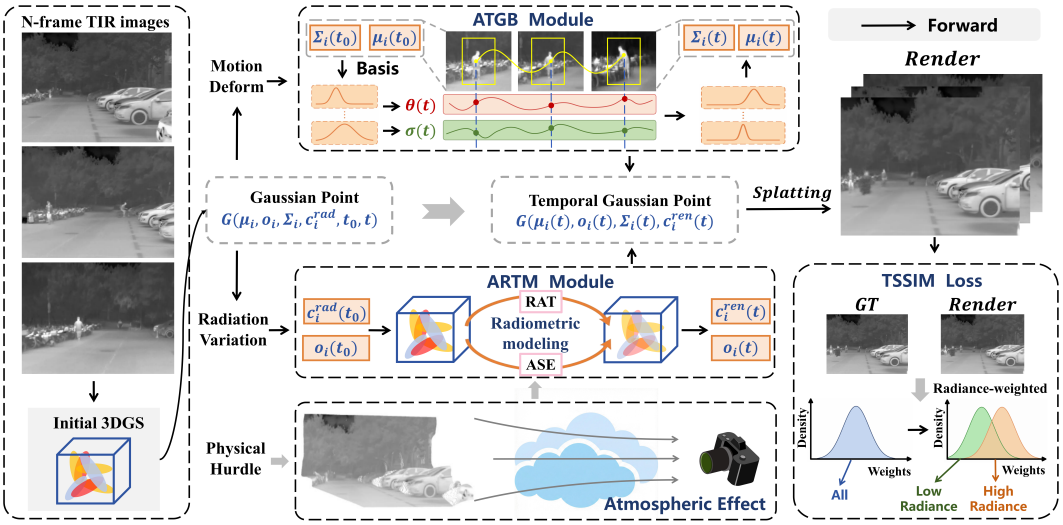


Figure 2: The overall framework of DTGS. DTGS decomposes the scene dynamics into two parts: Motion Deform and Radiation Variation, and designs the Adaptive Temporal Gaussian Basis (ATGB) and Atmospheric Radiative Transport Modeling (ARTM) Module to fit and predict the respective changes. Meanwhile, TSSIM loss can enhance network's focus on moving foreground targets.

Among them, D-NeRF [11] adopts a two-stage framework: it first encodes the scene into a canonical space, then maps this representation to a time-varying deformed scene. HexPlane [19] introduces six 2D feature plane combined with a lightweight MLP and volume rendering for dynamic novel view synthesis. However, these methods suffer from long training times and slow rendering speeds.

Recently, 3DGS [9] introduced a point-based rendering technique that simultaneously achieves high rendering quality and extremely fast speed, inspiring many Gaussian-based dynamic reconstruction methods [21, 22, 23, 24]. Among them, 4DGS [25] combines 3D Gaussians with 4D neural voxels and uses a lightweight MLP to decode geometric deformations, while LocalDyGS [26] models complex dynamics by decomposing scenes into local spaces and decoupling static and dynamic features. Although GS-based methods perform well in visible-light dynamic scenes, they often produce artifacts and blurring in thermal scenes due to fundamental differences in imaging mechanisms.

2.2 Thermal 3D perception

Motivated by the applications of NeRF and 3DGS in visible light scenes, several works [27, 28, 29, 30] have extended them to thermal infrared scenarios. Thermal3D-GS [30] leverages neural networks to model heat conduction in 3D media, enhancing thermal infrared reconstruction accuracy. NTR-Gaussian [31] further treats temperature as thermal radiation and incorporates convective heat transfer and radiative cooling to predict the temporal evolution of temperature in 3D scenes.

However, these methods either focus on static thermal scenes or only model temperature evolution. When a thermal infrared scene exhibits both complex geometric changes due to motion and varying radiation intensity, they struggle to achieve accurate reconstruction. To overcome this limitation, this paper proposes a physics-embedded dynamic thermal 3D reconstruction method, named as DTGS. By embedding the physics of thermal infrared radiation into the model, our approach can jointly predict both radiation intensity changes and dynamic motion.

3 Method

In this section, we first elaborate on the basic parameter settings in 3DGS. Next, we introduce the Atmospheric Radiative Transport Modeling (ARTM) module, which enables the dynamic computation of color and opacity parameters. Then, we present the Adaptive Temporal Gaussian Basis (ATGB) module, which facilitates the dynamic computation of position, scale, and rotation parameters. Finally, we discuss the TSSIM loss, which enhances the model's ability to focus on regions of interest.

3.1 3DGS Parameter Settings

Each 3DGS point at time t is jointly represented by a covariance matrix $\Sigma_i(t)$ and a center position $\mu_i(t) \in R^3$ ($i = 1, 2, 3...n$), where the covariance matrix represents the rotation and scaling parameters. In the thermal infrared scene, the radiance and opacity of each Gaussian point are denoted by $c_i^{\text{rad}}(t)$ and $o_i(t) \in [0, 1]$ respectively. Through geometric and radiometric modeling of these attributes, amount of changes at the current time t relative to the initial time can be obtained as:

$$(\Delta\mu_i(t), \Delta o_i(t), \Delta\Sigma_i(t), \Delta c_i^{\text{rad}}(t)) = F(\mu_i(t_0), o_i(t_0), \Sigma_i(t_0), c_i^{\text{rad}}(t_0), t), \tag{1}$$

where $\mu_i(t_0), o_i(t_0), \Sigma_i(t_0), c_i^{\text{rad}}(t_0)$ represent the position, opacity, covariance matrix, and radiance of the Gaussian point at the initial time t_0.

3.2 ARTM: Atmospheric Radiative Transport Modeling

Visible-light imaging and long-wave infrared (LWIR) imaging are governed by fundamentally different physical mechanisms. In the LWIR band (8–14 μm), scene radiance is dominated by objects' intrinsic thermal emission rather than reflected light. During propagation to the sensor, this radiation is strongly influenced by the atmosphere: gas molecules such as water vapor and CO_2 absorb infrared radiation, causing Radiation Attenuation (RAT), while the warm atmosphere itself emits additional thermal radiation, termed Atmospheric Spontaneous Emission Enhancement (ASE) [32]. In dynamic infrared scenes, these two effects directly modulate the sensor-measured radiance. Ignoring them in modeling leads to Gaussian radiance and opacity parameters deviations from the true physical properties of the scene.

Based on the above observations, we explicitly introduce atmospheric radiative transport into the 3D Gaussian Splatting rendering pipeline to establish a physically consistent imaging model from object-surface thermal emission to sensor measurements. Specifically, along the propagation path s from a scene point to the camera center, the spectral radiance $L_\lambda(s)$ at wavelength λ is described by the Schwarzschild [33] form of the radiative transfer equation:

$$\frac{dL_\lambda(s)}{ds} = -\beta_\lambda(s)L_\lambda(s) + \beta_\lambda(s)B_\lambda(T(s)), \tag{2}$$

where $\beta_\lambda(s)$ denotes the atmospheric extinction coefficient at position s, and $B_\lambda(T(s))$ is the Planck blackbody radiance at temperature $T(s)$. By integrating Eq. (2) along the propagation path, we obtain the spectral radiance at distance d:

$$L_\lambda(d) = L_\lambda(0)\exp\left(-\int_0^d \beta_\lambda(u)\,du\right) + \int_0^d \beta_\lambda(s)\,B_\lambda(T(s))\exp\left(-\int_s^d \beta_\lambda(u)\,du\right)\,ds, \tag{3}$$

the first term corresponds to the object-emitted radiance that reaches the sensor after atmospheric attenuation, whereas the second term captures the accumulated atmospheric self-emission along the propagation path, where the contribution from each infinitesimal path segment is attenuated by the remaining atmospheric medium between that segment and the sensor. For notational convenience, we define the optical thickness from the starting point to distance d as $\tau_\lambda(0, d) = \int_0^d \beta_\lambda(u)\,du$, and the residual optical thickness from an intermediate position s to the endpoint d as $\tau_\lambda(s, d) = \int_s^d \beta_\lambda(u)\,du$. Then Eq. (3) can be written as:

$$L_\lambda(d) = L_\lambda(0)\exp\big(-\tau_\lambda(0, d)\big) + \int_0^d \beta_\lambda(s)\,B_\lambda(T(s))\,\exp\big(-\tau_\lambda(s, d)\big)\,ds. \tag{4}$$

Within the long-wave infrared atmospheric window, the atmosphere exhibits relatively high transmittance to infrared radiation, and the contribution of atmospheric self-emission is typically smaller than that of target thermal emission. Moreover, in typical near-ground observation scenarios, the propagation distance is short, and the atmospheric self-emission term varies slowly across space, allowing it to be approximated as a pixel-wise uniform background radiance offset. Therefore, for computational convenience, we represent the effect of atmospheric self-emission in the above formulation using a bias correction term $L(t)$. Let $c_i^{\text{rad}}(t)$ denote the thermal radiance associated

with Gaussian point G_i at time t. Taking atmospheric attenuation into account, the effective radiance used for rendering $c_i^{\text{ren}}(t)$ is given by:

$$c_i^{\text{ren}}(t) = e^{-\tau_\lambda(0,d_i)}\, c_i^{\text{rad}}(t), \tag{5}$$

where d_i is the distance from the i-th Gaussian point to the camera center. After accounting for atmospheric attenuation and atmospheric self-emission, the radiance at pixel p and time t under the rasterization-based rendering formulation is given by:

$$C_p(t) = \sum_{i \in \mathcal{N}_p} c_i^{\text{ren}}(t)\, \alpha_i'(t) \prod_{j=1}^{i-1} \big(1 - \alpha_j'(t)\big) + L_p(t), \tag{6}$$

where $\alpha_i'(t)$ denotes effective opacity of the i-th Gaussian projected onto pixel p, which is computed as the product of the Gaussian response and the opacity parameter $o_i(t)$. Here, $L_p(t)$ denotes bias correction term for pixel p at time t. $\mathcal{N}_p$ denote the Gaussian ellipsoid that contributes to pixel p.

To capture temporal variations in dynamic scenes, we model the opacity $o_i(t)$, the atmospheric attenuation factor $\eta_i(t) = e^{-\tau_\lambda(0,d_i)}$, the bias term $L_p(t)$ as time-varying quantities. In ARTM module, each of these scalars is approximated by a weighted sum of static temporal Gaussian basis functions along the time axis:

$$\tilde{b}(t;\theta,\sigma) = \exp\left(-\frac{1}{2\sigma^2}(t-\theta)^2\right), \tag{7}$$

where t denotes time, θ is the center of the Gaussian basis, and σ controls its spread. Accordingly, a generic time-varying quantity can be approximated by a weighted sum of Gaussian bases. Taking the atmospheric attenuation factor $\eta_i(t)$ as an example, we write

$$\psi^{\eta_i(t)}\left(t, \xi^{\eta_i(t)}\right) = \sum_{j=1}^{B} \gamma_j^{\eta_i(t)}\, \tilde{b}\left(t; \theta_j^{\eta_i(t)}, \sigma_j^{\eta_i(t)}\right), \tag{8}$$

where $\xi^{\eta_i(t)} = \left\{\gamma_j^{\eta_i(t)}, \theta_j^{\eta_i(t)}, \sigma_j^{\eta_i(t)}\right\}$ are learnable parameters. Here, $\gamma_j^{\eta_i(t)}$ is the weight of the j-th Gaussian basis, j indexes the basis functions, and B is total number of bases, which is taken as 15.

3.3 ATGB: Adaptive Temporal Gaussian Basis

In dynamic thermal infrared scenes, moving objects (e.g., pedestrians and vehicles) act as endogenous heat sources with continuously changing geometry. ARTM module uses **static temporal Gaussian bases** to approximate thermal radiation variation processes, but such fixed bases (with time-invariant γ, θ, and σ) are insufficient to model complex, non-stationary motion. In ATGB module, we therefore introduce an **Adaptive Temporal Gaussian basis** for motion geometric parameters (positions, scales, rotations), where the temporal mean θ and variance σ are allowed to evolve over time under simple, physics-inspired kinematic priors.

The trajectory of a dynamic heat source can typically be decomposed into a globally smooth, approximately linear translation and a local periodic component (e.g., wheel rotation or limb swing in human gait). Motivated by this observation, we let the temporal center $\theta_j(t)$ of the j-th Gaussian basis evolve along the time axis as the sum of a linear drift and a sinusoidal oscillation:

$$\theta_j(t) = \theta_j(t_0) + v_{\theta_j}\, t + \Delta\theta_j\, \sin\big(\omega_{\theta_j}\, t + \phi_{\theta_j}\big), \tag{9}$$

where $\theta_j(t_0)$ is the initial temporal anchor of the j-th basis, v_{θ_j} is the drift rate of this anchor along the time axis, and $\Delta\theta_j$, ω_{θ_j}, ϕ_{θ_j} denote the amplitude, angular frequency, and phase of the periodic component. The linear term $v_{\theta_j} \cdot t$ shifts the activation center of each basis to follow the macroscopic translation of the observed object, while the sinusoidal term allows the center to oscillate in accordance with fine-grained periodic rhythms. This adaptive temporal parameterization substantially improves the smoothness of the model in predicting thermal infrared target dynamics.

The temporal variance σ of each basis function controls the effective duration over which the basis contributes to the fitting. For static or slowly moving objects a fixed σ suffices, but for dynamic heat sources the appropriate temporal support varies over time: an object undergoing acceleration,

depth-induced scale change, or thermal boundary diffusion requires a wider or narrower temporal window at different stages. Motivated by these observations, we allow σ to evolve over time and adopt an exponential parameterization to guarantee positivity:

$$\sigma_j(t) = \sigma_j(t_0) \cdot \exp\Big(v_{\sigma_j}\, t + \Delta\sigma_j \cdot \sin\big(\omega_{\sigma_j} \cdot t + \phi_{\sigma_j}\big)\Big). \tag{10}$$

where $\sigma_j(t_0)$ is the initial temporal scale, v_{σ_j} controls its drift rate along the time axis, and $\Delta\sigma_j$, ω_{σ_j}, ϕ_{σ_j} denote the amplitude, angular frequency, and phase of the periodic component. The exponential drift mirrors the slowly varying tendency of the apparent radiative extent (e.g., due to depth motion or thermal diffusion), while the sinusoidal term captures periodic fluctuations arising when different heated surfaces are alternately exposed under the viewing geometry.

Based on the temporally evolving mean $\theta_j(t)$ (Eq. (9)) and variance $\sigma_j(t)$ (Eq. (10)), the adaptive temporal Gaussian basis function is defined as:

$$\widetilde{tb}_e\big(t;\ \theta_j(t),\ \sigma_j(t)\big) = \exp\left(-\frac{\big(t-\theta_j(t)\big)^2}{2\,\sigma_j(t)^2}\right) \tag{11}$$

Note that $\widetilde{tb}_e$ remains an 1D Gaussian kernel along the time axis; only its parameters $\theta_j(t)$ and $\sigma_j(t)$ are time-dependent, so the argument $(t-\theta_j(t))$ is always scalar. Taking the position $\mu_i(t)$ as an example, the deformation at time t is approximated by a weighted sum of coupled bases:

$$\varphi^{\mu_i(t)}(t) = \sum_{j=1}^{B} \gamma_j^{\mu_i(t)}\ \widetilde{tb}_e\Big(t;\ \theta_j^{\mu_i(t)}(t),\ \sigma_j^{\mu_i(t)}(t)\Big) \tag{12}$$

where $\gamma_j^{\mu_i(t)}$ is the weight of the j-th basis and B is the total number of bases, which is taken as 15. Although each individual basis adopts a simple linear-plus-sinusoidal parameterization, the local envelope property of $\widetilde{tb}_e$, together with the weighted superposition of multiple bases, endows the model with the expressiveness to approximate not only periodic rhythmic patterns but also arbitrary aperiodic transient motions (e.g., sudden acceleration or irregular thermal diffusion) within a finite temporal domain through multi-frequency superposition.

3.4 TSSIM Loss

Considering that the standard SSIM loss does not adapt well to the physical characteristics of infrared images, by analyzing Planck's radiation law [34] and the signal-to-noise characteristics of thermal radiation [35], we design the TSSIM loss. This loss enhances the model's ability to focus on dynamic objects in the scene while reducing the interference caused by noise.

According to Planck's law of blackbody radiation, the spectral radiance of an object at wavelength λ is given by

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \cdot \frac{1}{\exp\Big(\frac{hc}{\lambda k_B T}\Big) - 1}, \tag{13}$$

where h denotes Planck's constant, c is the speed of light, k_B is the Boltzmann constant, and T is the object's absolute temperature. In the long-wave infrared (LWIR) band, this relation implies that higher temperatures yield stronger thermal emission, observed as higher pixel intensities in infrared images. In typical road and outdoor scenes, moving targets such as vehicles and pedestrians generally have higher surface temperatures than the surrounding static background (e.g., buildings and pavement) due to metabolic or mechanical heat generation. They therefore appear as bright, high-emissive regions and constitute the main objects of interest for scene understanding and reconstruction. However, the standard SSIM loss uses isotropic weighting over the image, preventing the model from assigning greater importance to these high-emissive regions that carry richer semantic information.

Meanwhile, the signal-to-noise ratio (SNR) of an infrared focal-plane array detector is related to the target radiance intensity I as: $\mathrm{SNR} \propto \sqrt{I}$. The detector noise is dominated by photon shot noise,

denoted as σ_{noise}, which follows a similar dependence on the radiance intensity: $\sigma_{\text{noise}} \propto \sqrt{I}$. These relations indicate that in high-radiance regions (e.g., moving vehicles and pedestrians), the structural information in the image is relatively reliable. In contrast, in background regions (e.g., buildings and road surfaces) where the radiance intensity is low, pixel values are more severely corrupted by noise. However, the conventional SSIM metric assigns equal weights to all pixels, causing model fitting to be disproportionately affected by noise in low-temperature (low-radiance) regions.

Based on the physical mechanisms described above, this paper proposes a thermal radiation intensity-weighted structural similarity loss function, referred to as TSSIM. By employing the radiation intensity I_i of each pixel in the ground-truth image as an adaptive weight, thermal reweighting is applied to the image statistics, including the mean, variance, and covariance. This design automatically shifts the optimization focus of the loss function toward high-temperature, high-SNR key target regions, while suppressing the gradient interference caused by low-temperature noisy regions.

Taking the rendered image X and the ground-truth image Y as an example, the thermally weighted mean μ_X^{th}, variance $(\sigma_X^{th})^2$, and covariance σ_{XY}^{th} are computed as follows:

$$\mu_X^{\text{th}} = \frac{\sum_i I_i x_i}{\sum_i I_i}, \quad (\sigma_X^{\text{th}})^2 = \frac{\sum_i I_i (x_i - \mu_X^{\text{th}})^2}{\sum_i I_i}, \quad \sigma_{XY}^{\text{th}} = \frac{\sum_i I_i (x_i - \mu_X^{\text{th}})(y_i - \mu_Y^{\text{th}})}{\sum_i I_i} \tag{14}$$

The final expression for the TSSIM function is therefore given by:

$$L_{\text{TSSIM}}(X, Y) = \frac{(2\mu_X^{th}\mu_Y^{th} + C_1)(2\sigma_{XY}^{th} + C_2)}{((\mu_X^{th})^2 + (\mu_Y^{th})^2 + C_1)((\sigma_X^{th})^2 + (\sigma_Y^{th})^2 + C_2)}. \tag{15}$$

In the above formulation, C_1 and C_2 are set to the same stabilizing constants as those used in the standard SSIM loss. The total loss function of the system is the sum of L_1 color loss and L_{TSSIM}, which can be expressed as:

$$L_{\text{total}} = 0.7 \cdot L_1 + 0.3 \cdot (1 - L_{\text{TSSIM}}). \tag{16}$$

4 Experiments

This section proceeds as follows: we first present the proposed Dynamic_LTR dataset in Section 4.1, followed by the training protocols and hyperparameter configurations in Section 4.2. Comprehensive comparative benchmarks and ablation studies on Dynamic_LTR are detailed in Sections 4.3 and 4.4.

4.1 Dynamic_LTR Dataset

We capture and collect a thermal infrared dynamic 3D scene reconstruction dataset, Dynamic_LTR. Dynamic_LTR comprises 31 scenes spanning diverse low-visibility conditions (nighttime, overcast, hazy) where RGB sensing is unreliable; 26 of these scenes were captured by our team, and the remaining 5 were selected from the ZJU-Multispectral dataset [36] as they align with the topic of this paper. Following the dataset preprocessing protocols adopted in DyNeRF [11] and TI-NSD [37], we preprocess all scenes in Dynamic_LTR. Based on the number of moving dynamic objects present in each scene, as well as the indoor versus outdoor nature of the scenes, we further partition Dynamic_LTR into four subsets: SOD_indoor (including 4 scenes), SOD_outdoor (including 4 scenes), MOID_indoor (including 3 scenes) and MOID_outdoor (including 20 scenes). Here, SOD denotes Single-Object Dynamic, while MOID represents Multi-Object Interactive Dynamic. A detailed description of the dataset, along with a comparison to other datasets, can be found in **Appendix A.5**.

4.2 Training Details

Our implementation is primarily based on PyTorch framework and tested on a single RTX 3090 GPU. The training involved a total of 3500 iterations, with optimization performed using a single Adam optimizer. As for training, we partition the data into training and test sets, every 8th image is selected for the test set to ensure effective and meaningful comparisons for generating error metrics.

Table 1: Reconstruction Accuracy Comparison on the Dynamic_LTR Dataset

Method	SOD_indoor			SOD_outdoor			MOID_indoor			MOID_outdoor		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
4DGS [25]	28.34	0.895	0.298	26.96	0.812	0.306	22.52	0.843	0.349	27.65	0.877	0.301
LocalDyGS [26]	26.72	0.881	0.335	21.14	0.751	0.343	22.17	0.822	0.379	23.72	0.814	0.374
RD4DGS [38]	31.16	0.908	0.280	<u>28.14</u>	0.831	0.261	<u>24.75</u>	0.844	0.322	28.11	0.873	0.292
NTR-Gaussian [31]	25.24	0.864	0.347	22.89	0.790	0.342	21.31	0.805	0.415	24.39	0.828	0.366
Thermal3D-GS [30]	<u>31.80</u>	<u>0.925</u>	<u>0.247</u>	27.81	<u>0.874</u>	<u>0.218</u>	24.43	<u>0.872</u>	<u>0.285</u>	<u>28.76</u>	<u>0.900</u>	<u>0.251</u>
DTGS (Ours)	**34.26**	**0.953**	**0.191**	**30.58**	**0.926**	**0.175**	**26.62**	**0.903**	**0.203**	**31.59**	**0.934**	**0.173**

Meanwhile, we employ a variety of metrics to evaluate the experimental results, including PSNR, LPIPS and SSIM. For comparison, we select three dynamic reconstruction methods in visible light scenarios: 4DGS [25], LocalDyGS [26], and RD4DGS [38], as well as two models with dynamic reconstruction capabilities in thermal scenarios: NTR-Gaussian [31] and Thermal3D-GS [30].

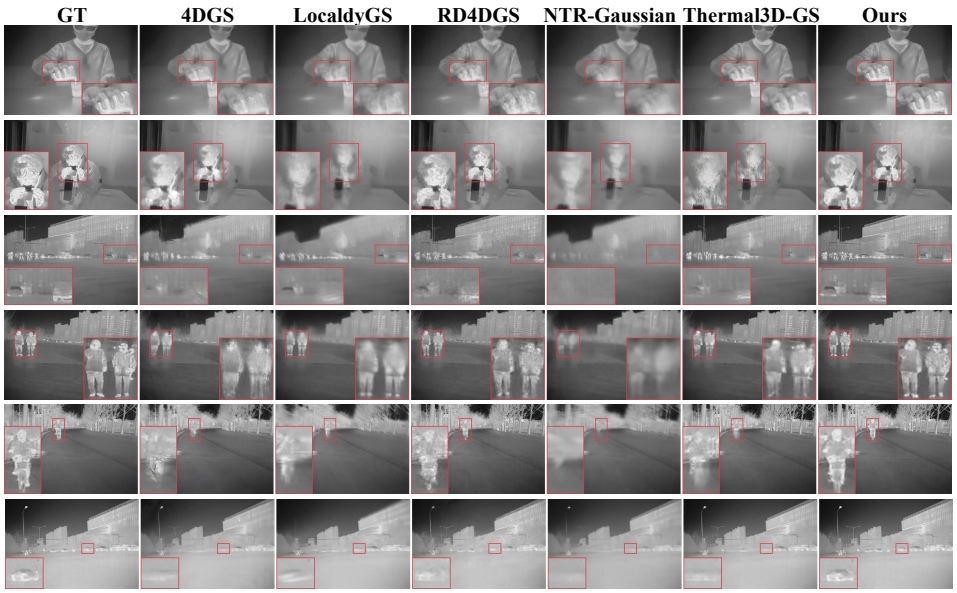


Figure 3: Comparative visualization results on the Dynamic_LTR dataset

4.3 Evaluation on Dynamic_LTR Dataset

As reported in Table 1, Our method achieves PSNR improvements of 2.46 dB, 2.44 dB, 1.87 dB, and 2.83 dB on the four sub-datasets respectively. The qualitative results in Fig. 3 further validate its effectiveness on the Dynamic_LTR dataset. NTR-Gaussian produces heavily blurred reconstructions, as it only models temporal temperature variations but ignores geometric deformation. Thermal3D-GS exhibits a certain degree of dynamic modeling and predictive capability but still fails to recover complex structural changes, leading to noticeable blur under large motions. In contrast, visible-light dynamic reconstruction methods (4DGS, LocalDyGS, RD4DGS) also suffer from artifacts and blurring in their renderings due to differences in the imaging mechanisms between visible light and infrared. For instance, in the first row of comparisons, the reconstruction of the hand exhibits blending with the cup. In the sixth row, depicting a road traffic scene, all three aforementioned methods fail to accurately compute the trajectories of fast-moving vehicles. Fig. 4 presents a visual comparison of the dynamic reconstruction of the same scene at different time instants. The results indicate that, even in challenging cases where previously observed dynamic objects disappear and new dynamic targets emerge, our method is still able to accurately fit the motion trajectories of multiple dynamic objects.

Table 2: Comprehensive Metrics Comparison on the Dynamic_LTR Dataset

Method	Tr_iter	Avg_Tr_time(s)↓	Avg_FPS↑	Avg_PSNR↑	Avg_SSIM↑	Avg_LPIPS↓
4DGS [25]	20000	1595	30	27.15	0.868	0.306
LocalDyGS [26]	20000	848	58	23.62	0.815	0.365
RD4DGS [38]	10000	524	67	28.18	0.869	0.289
NTR-Gaussian [31]	15000	778	32	24.01	0.826	0.365
Thermal3D-GS [30]	15000	720	36	28.61	0.897	0.250
DTGS (Ours)	3500	**173**	**127**	**31.32**	**0.932**	**0.178**

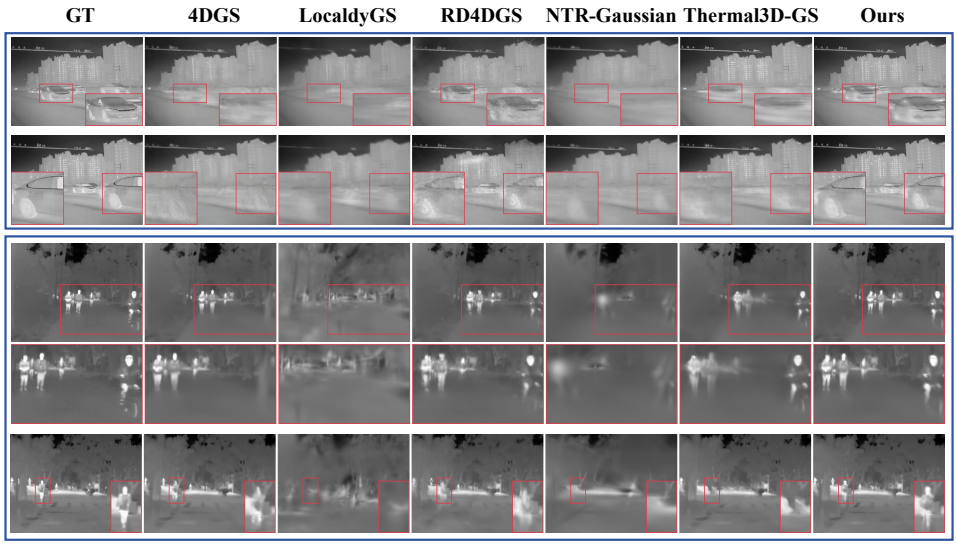


Figure 4: Comparative visual comparison of the same scene at different time instants.

Table 2 presents the training iterations for convergence, average training time, inference speed, and reconstruction accuracy of six methods across 31 scenes in the Dynamic_LTR dataset. Experimental results demonstrate that, in terms of reconstruction accuracy across all scenes, our method outperforms current state-of-the-art approaches by 9.5%, 3.9%, and 28.8% in PSNR, SSIM, and LPIPS, respectively. These improvements stem from the ARTM and TSSIM modules in DTGS, which are specifically designed to account for the physical characteristics of infrared imaging, effectively addressing the challenges posed by infrared-specific physical dilemmas in reconstruction. Furthermore, the ATGB module employs Adaptive Temporal Gaussian basis functions, enabling accurate fitting of complex motion trajectories within the scene. In addition, DTGS achieves a 3× speedup in training time and a 2× speedup in inference time compared to current state-of-the-art methods. This efficiency gain is attributed to the absence of deep learning or MLP modules in the DTGS framework; instead, it relies on basis functions to model dynamic changes, which significantly accelerates both training and inference. More comparative experimental results can be found in **Appendix A.1**.

4.4 Ablation study

Table 3 presents ablation visualizations of three proposed modules on Dynamic_LTR dataset. To highlight the capability of the proposed modules to address the challenges that complex motions and physical properties impose on dynamic reconstruction, we compare the ATGB module with static Gaussian_basis functions, and we further compare the proposed TSSIM loss with the conventional SSIM loss. The results show that incorporating ARTM and ATGB improves the reconstruction accuracy by 1.05 dB and 1.54 dB in terms of average PSNR, which substantiates the effectiveness of the proposed method. In addition, TSSIM achieves a 0.63 dB gain over SSIM, indicating that TSSIM better emphasizes the reconstruction quality of dynamic physical variations. The ablation visualization results for each individual module can be found from **Appendix A.2 to A.4**.

Table 3: Ablation experiments on the Dynamic_LTR Dataset

ARTM	Gaussian_basis	ATGB	SSIM	TSSIM	Avg_PSNR↑	Avg_SSIM↑	Avg_LPIPS↓
		✓		✓	30.27	0.923	0.203
✓	✓			✓	29.78	0.914	0.216
✓		✓	✓		30.69	0.920	0.245
✓		✓		✓	**31.32**	**0.932**	**0.178**

5 Conclusion

In this paper, we propose a physics-embedded dynamic thermal infrared 3D reconstruction method, termed DTGS. DTGS incorporates thermal infrared physical characteristics into the 3DGS model architecture, effectively mitigating the impact of atmospheric transmission effects and noise interference in low signal-to-noise background regions. In addition, we introduce an adaptive temporal Gaussian basis to model motion deformations. Ultimately, DTGS is capable of fitting and predicting the dynamics within the scene. To evaluate the effectiveness of the proposed approach, we construct a large-scale dataset Dynamic_LTR. Experimental results demonstrate that DTGS achieves substantial improvements in reconstruction accuracy, training efficiency, and rendering speed.

Acknowledgements

This work is supported by Beijing Natural Science Foundation under Grant L258070, National Natural Science Foundation of China under Grant 62271016 and the Shenzhen Natural Science Foundation JCYJ20230807112620040.

Competing Interests

The authors declare no competing financial or non-financial interests.

References

[1] Chaoqun Wang, Xiangyu Chen, Chenming Li, Rui Song, Yibin Li, and Max Q-H Meng. Chase and track: Toward safe and smooth trajectory planning for robotic navigation in dynamic environments. *IEEE Transactions on Industrial Electronics*, 70(1):604–613, 2022.

[2] Junbo Li, Weimin Yuan, Yinuo Wang, Yue Zeng, Shihao Shu, Cai Meng, and Xiangzhi Bai. Gs2pose: Marry gaussian splatting to 6d object pose estimation. *arXiv preprint arXiv:2510.16777*, 2025.

[3] Junbo Li, Keyan Chen, Gengju Tian, Lu Li, and Zhenwei Shi. Marsseg: Mars surface semantic segmentation with multilevel extractor and connector. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–12, 2025.

[4] Jing Yang, Qinghua Ni, Guiyang Luo, Qi Cheng, Latifa Oukhellou, and Shuangshuang Han. A trustworthy internet of vehicles: The dao to safe, secure, and collaborative autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 8(12):4678–4681, 2023.

[5] Lorenzo Barruffo, Bianca Caiazzo, Alberto Petrillo, and Stefania Santini. A goa4 control architecture for the autonomous driving of high-speed trains over etcs: Design and experimental validation. *IEEE Transactions on Intelligent Transportation Systems*, 25(6):5096–5111, 2024.

[6] Udaya LN Puvvadi, Kevin Di Benedetto, Aditya Patil, Kyoung-Don Kang, and Youngjoon Park. Cost-effective security support in real-time video surveillance. *IEEE Transactions on Industrial Informatics*, 11(6):1457–1465, 2015.

[7] Xin Yuan, Shuyan Hu, Wei Ni, Xin Wang, and Abbas Jamalipour. Deep reinforcement learning-driven reconfigurable intelligent surface-assisted radio surveillance with a fixed-wing uav. *IEEE Transactions on Information Forensics and Security*, 18:4546–4560, 2023.

[8] 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.

[9] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, et al. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

[10] Junoh Lee, ChangYeon Won, Hyunjun Jung, Inhwan Bae, and Hae-Gon Jeon. Fully explicit dynamic gaussian splatting. *Advances in Neural Information Processing Systems*, 37:5384–5409, 2024.

[11] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10318–10327, 2021.

[12] Yifan Zhan, Zhuoxiao Li, Muyao Niu, Zhihang Zhong, Shohei Nobuhara, Ko Nishino, and Yinqiang Zheng. Kfd-nerf: Rethinking dynamic nerf with kalman filter. In *European Conference on Computer Vision*, pages 1–18. Springer, 2024.

[13] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20331–20341, 2024.

[14] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5712–5721, 2021.

[15] Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12882–12891, 2022.

[16] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228*, 2021.

[17] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6498–6508, 2021.

[18] Xunzhi Zheng and Dan Xu. Flow-nerf: Joint learning of geometry, poses, and dense flow within unified neural representations. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 993–1002, 2025.

[19] Ang Cao and Justin Johnson. Hexplane: A fast representation for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 130–141, 2023.

[20] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12479–12488, 2023.

[21] Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, and Wei Xing. 3dgstream: On-the-fly training of 3d gaussians for efficient streaming of photo-realistic free-viewpoint videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20675–20685, 2024.

[22] Jiahui Lei, Yijia Weng, Adam W Harley, Leonidas Guibas, and Kostas Daniilidis. Mosca: Dynamic gaussian fusion from casual videos via 4d motion scaffolds. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 6165–6177, 2025.

[23] Youtian Lin, Zuozhuo Dai, Siyu Zhu, and Yao Yao. Gaussian-flow: 4d reconstruction with dynamic 3d gaussian particle. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21136–21145, 2024.

[24] Ruijie Zhu, Yanzhe Liang, Hanzhi Chang, Jiacheng Deng, Jiahao Lu, Wenfei Yang, Tianzhu Zhang, and Yongdong Zhang. Motiongs: Exploring explicit motion guidance for deformable 3d gaussian splatting. *Advances in Neural Information Processing Systems*, 37:101790–101817, 2024.

[25] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 20310–20320, 2024.

[26] Jiahao Wu, Rui Peng, Jianbo Jiao, Jiayu Yang, Luyang Tang, Kaiqiang Xiong, Jie Liang, Jinbo Yan, Runling Liu, and Ronggang Wang. Localdygs: Multi-view global dynamic scene modeling via adaptive local implicit feature decoupling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9519–9529, 2025.

[27] 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.

[28] Mariam Hassan, Florent Forest, Olga Fink, and Malcolm Mielle. Thermonerf: A multimodal neural radiance field for joint rgb-thermal novel view synthesis of building facades. *Advanced Engineering Informatics*, 65:103345, 2025.

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

[30] Qian Chen, Shihao Shu, Heng Sun, Junzhang Chen, and Xiangzhi Bai. Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis with a large-scale dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

[31] Kun Yang, Yuxiang Liu, Zeyu Cui, Yu Liu, Maojun Zhang, Shen Yan, and Qing Wang. Ntr-gaussian: Nighttime dynamic thermal reconstruction with 4d gaussian splatting based on thermodynamics. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 691–700, 2025.

[32] Tiangang Yin, Simone Kotthaus, Jean-Philippe Gastellu-Etchegorry, William Morrison, Leslie K Norford, Sue Grimmond, Nicolas Lauret, Nektarios Chrysoulakis, Lucas Landier, and Ahmad Al Bitar. Atmospheric correction of ground-based thermal infrared camera through dart model. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5885–5888. IEEE, 2017.

[33] Karl Schwarzschild. Über das gleichgewicht der sonnenatmosphäre. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1906:41–53, 1906.

[34] Max Planck. 2. zur theorie des gesetzes der energieverteilung im normalspektrum. In *Quantentheorie*, pages 107–117. De Gruyter, 2022.

[35] Z Bielecki. Maximisation of signal-to-noise ratio in infrared radiation receivers. *OPTOELECTRONICS REVIEW*, (3):209–216, 2002.

[36] Weichen Dai, Yu Zhang, Shenzhou Chen, Donglei Sun, and Da Kong. A multi-spectral dataset for evaluating motion estimation systems. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5560–5566. IEEE, 2021.

[37] 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.

[38] Hyeongmin Lee and Kyungjune Baek. Temporal smoothness-aware rate-distortion optimized 4d gaussian splatting. *arXiv preprint arXiv:2507.17336*, 2025.

A Technical appendices and supplementary material

A.1 Comparative Experiments on Dynamic_LTR Dataset

Table 4 and Table 5 present the per-scene accuracy comparison between our method and current state-of-the-art methods across four sub-datasets: SOD_indoor, SOD_outdoor, MOID_indoor, and MOID_outdoor. Experimental results show that DTGS achieves the highest dynamic reconstruction accuracy in 13 out of 14 scene categories in terms of the PSNR metric, while attaining the best performance across all categories under the SSIM and LPIPS metrics. These results further validate the effectiveness of our method for thermal infrared dynamic scene reconstruction.

Table 4: Per-category results on SOD_indoor and SOD_ outdoor subsets.

Metric	Method	eating	Tai Chi	Pouring water	Dribble	Layup	Shooting	Traffic single	Avg.
PSNR↑	Thermal3DGS	30.47	25.41	35.51	26.26	25.70	26.89	32.40	28.95
	4DGS	27.86	22.56	30.85	24.79	22.37	24.01	**36.67**	27.02
	NTR-Gaussian	25.44	20.36	26.69	22.52	19.29	23.16	26.60	23.44
	LocalDyGS	25.78	21.42	29.51	23.41	21.43	21.91	17.81	23.04
	RD4DGS	30.50	25.43	33.78	25.72	25.32	26.33	35.17	28.89
	DTGS(Ours)	**33.71**	**28.58**	**37.32**	**28.04**	**28.51**	**29.21**	36.56	**31.70**
SSIM↑	Thermal3DGS	0.931	0.862	0.951	0.852	0.819	0.862	0.961	0.891
	4DGS	0.912	0.823	0.920	0.806	0.698	0.774	0.971	0.843
	NTR-Gaussian	0.873	0.780	0.899	0.758	0.674	0.720	0.920	0.803
	LocalDyGS	0.898	0.802	0.907	0.777	0.751	0.736	0.742	0.802
	RD4DGS	0.914	0.838	0.934	0.780	0.775	0.799	0.968	0.858
	DTGS(Ours)	**0.957**	**0.913**	**0.972**	**0.899**	**0.899**	**0.928**	**0.976**	**0.935**
LPIPS↓	Thermal3DGS	0.247	0.284	0.232	0.230	0.298	0.173	0.172	0.234
	4DGS	0.279	0.339	0.289	0.307	0.439	0.324	0.153	0.304
	NTR-Gaussian	0.349	0.393	0.319	0.346	0.440	0.314	0.269	0.347
	LocalDyGS	0.356	0.371	0.308	0.333	0.339	0.332	0.368	0.344
	RD4DGS	0.279	0.334	0.265	0.318	0.322	0.253	0.150	0.274
	DTGS(Ours)	**0.206**	**0.228**	**0.169**	**0.202**	**0.210**	**0.150**	**0.137**	**0.186**

Table 5: Per-category results on MOID_indoor and MOID_ outdoor subsets

Metric	Method	Badminton	Cycling	Traffic	School dismissal	Escalator riding	Strolling	Mall Crowd	Avg.
PSNR↑	Thermal3DGS	24.43	27.03	30.05	27.13	26.51	28.20	27.33	27.24
	4DGS	22.52	25.41	29.59	25.92	24.11	25.87	24.71	25.45
	NTR-Gaussian	21.31	23.36	25.25	24.43	22.41	23.77	22.45	23.28
	LocalDyGS	22.17	24.29	23.36	25.14	23.18	24.56	23.30	23.71
	RD4DGS	24.75	26.84	29.21	27.05	28.02	28.35	24.96	27.02
	DTGS(Ours)	**26.62**	**29.70**	**32.73**	**29.38**	**29.46**	**30.76**	**31.15**	**29.97**
SSIM↑	Thermal3DGS	0.872	0.858	0.918	0.837	0.908	0.893	0.928	0.888
	4DGS	0.843	0.824	0.904	0.797	0.874	0.863	0.897	0.857
	NTR-Gaussian	0.805	0.770	0.854	0.736	0.854	0.832	0.845	0.814
	LocalDyGS	0.822	0.793	0.815	0.768	0.863	0.841	0.844	0.821
	RD4DGS	0.844	0.826	0.895	0.800	0.891	0.865	0.890	0.859
	DTGS(Ours)	**0.903**	**0.907**	**0.942**	**0.905**	**0.920**	**0.931**	**0.958**	**0.924**
LPIPS↓	Thermal3DGS	0.285	0.267	0.246	0.242	0.289	0.295	0.222	0.264
	4DGS	0.349	0.337	0.280	0.321	0.326	0.355	0.302	0.324
	NTR-Gaussian	0.415	0.375	0.361	0.379	0.345	0.407	0.361	0.377
	LocalDyGS	0.379	0.355	0.384	0.336	0.350	0.400	0.385	0.370
	RD4DGS	0.322	0.303	0.282	0.305	0.306	0.344	0.292	0.308
	DTGS(Ours)	**0.203**	**0.162**	**0.184**	**0.128**	**0.183**	**0.203**	**0.159**	**0.175**

Fig. 5 presents a comparison of dynamic predictions in more challenging scenes. Experimental results demonstrate that when objects in the scene undergo fast or irregular motion, both LocalDyGS and NTR-Gaussian suffer from reconstruction disorganization or severe prediction blurring. Although 4DGS and Thermal3D-GS exhibit some capability to model object motion in infrared dynamic

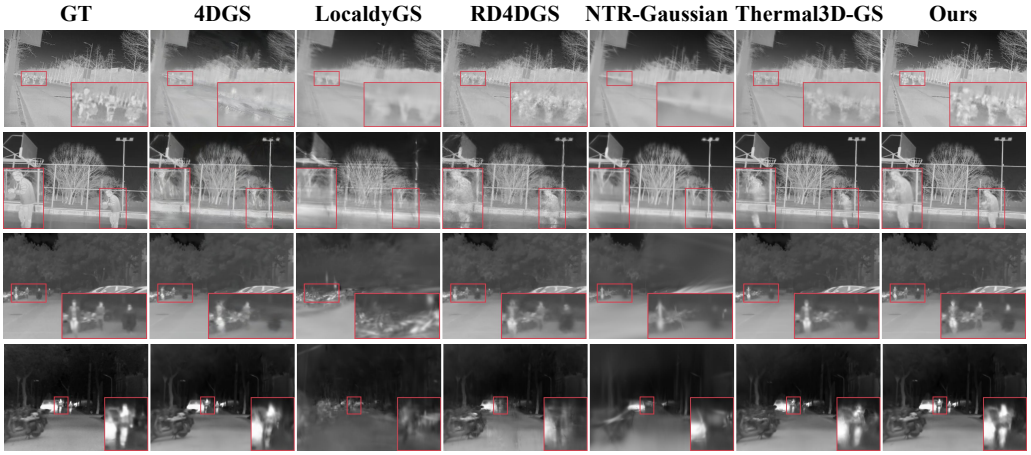


Figure 5: More comparative visualization results on the Dynamic_LTR dataset

scenes, they still encounter issues such as missing object predictions or significant loss of geometric structure. RD4DGS can roughly recover object motion patterns, but its predictions suffer from blurred boundaries and feature loss. In contrast, our method can reconstruct dynamic objects in the scene with higher accuracy while maintaining relatively clear boundaries and textures.

A.2 Visual Analysis of the ARTM Module

Fig. 6 presents our ablation visualization analysis for the ARTM module function. Experimental results demonstrate that the addition of the ARTM module improves the accuracy of reconstructing both geometric structures and textures of dynamic objects in the scene. In the first row of images, the ARTM module enhances the reconstruction accuracy of the contours of moving persons. In the second row of images, it improves the texture details on object surfaces in the reconstructed images.

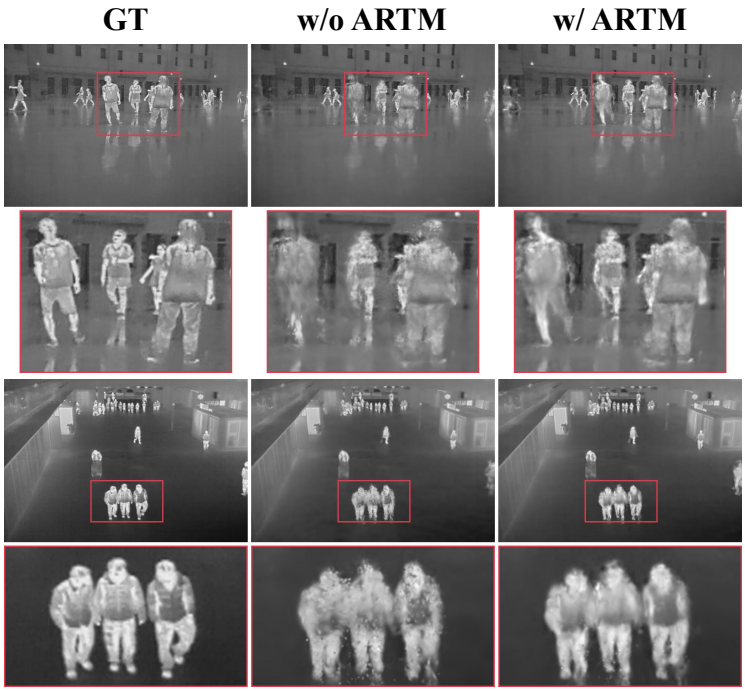


Figure 6: Ablation visualization of the ARTM module

A.3 Visual Analysis of the ATGB Module

Fig. 7 presents the ablation comparison visualization for the ATGB module. Experimental results demonstrate that, compared to Gaussian basis functions with fixed means and variances, the ATGB module exhibits a stronger capability to fit complex and fast motions. It yields fewer reconstruction artifacts for cyclists and moving vehicles, indicating that the model can more accurately predict the positions of dynamic objects.

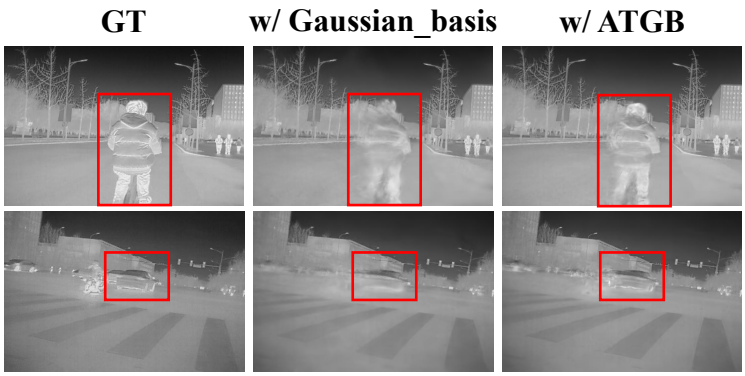


Figure 7: Ablation visualization of the ATGB module

A.4 Visual Analysis of the TSSIM Loss

Fig. 8 presents the ablation comparison visualization for the TSSIM loss. Compared to the SSIM loss, the TSSIM loss enhances the model's attention to dynamic endogenous heat sources with stronger radiation intensity in the scene, thereby improving the reconstruction accuracy of dynamic objects. The experimental visualization results show that using the TSSIM loss leads to higher precision in reconstructing both the overall contours and textures of dynamic objects.



Figure 8: Ablation visualization of the TSSIM loss

A.5 The Detailed Information of Dynamic_LTR Dataset

The Dynamic_LTR dataset comprises a total of 31 scenes, of which 26 are captured by our team, and the remaining 5 are selected from the ZJU-Multispectral dataset as they align with the topic of this paper. For data collection, we use a Zenmuse XT thermal camera equipped with an uncooled vanadium oxide microbolometer. During capture, the camera is either slowly moved by hand or kept stationary. Depending on the motion complexity of each scene, the recording duration ranges from 30 to 60 seconds, and approximately 100–150 frames are ultimately selected to form a scene segment.

Meanwhile, the 31 scenes are classified according to their scene types, resulting in a total of 14 scene categories. The number of scenes corresponding to each category is shown in Table 6.

Furthermore, based on the number of moving dynamic objects present in each scene, as well as the indoor versus outdoor nature of the scenes, we partition Dynamic_LTR into four subsets: SOD_indoor (include Eating, Tai_chi, Pouring_water), SOD_outdoor (including Dribble, Layup, Shooting, Traffic_single), MOID_indoor (include Badminton) and MOID_outdoor (include Cycling, School_dismissal, Strolling, Traffic, Mall_Crowd, Escalator_riding). Here, SOD denotes Single-Object Dynamic, while MOID represents Multi-Object Interactive Dynamic.

Table 6: Distribution of the Dynamic_LTR dataset

Category	Eating	Tai_chi	Pouring_water	Dribble	Layup	Shooting	Traffic_single
Number	1	1	2	1	1	1	1
Category	Badminton	Cycling	School_dismissal	Strolling	Traffic	Mall_Crowd	Escalator_riding
Number	3	3	2	1	11	2	1

For data preprocessing, we follow the approach used in DyNeRF and the TI-NSD dataset preprocessing pipeline, employing computational photography to obtain the scene point cloud and camera poses for each frame, which are then used as training input.

Table 7: A detailed comparison of Dynamic 3D reconstruction dataset

Dataset	Environment_Type	Total_number	Dynamic_Type
D-NeRF	RGB	8	Geometric Deformation
N3DV	RGB	6	Geometric Deformation
RHD	RGB & Thermal	10	Radiometric Deformation
TSDN	RGB & Thermal	5	Radiometric Deformation
NTR	Thermal	4	Radiometric Deformation
Dynamic_LTR	Thermal	31	Geometric & Radiometric Deformation

In Table 7, we present a comparison between the Dynamic_LTR dataset and current mainstream dynamic scene reconstruction datasets in thermal infrared and visible light settings. The Dynamic_LTR dataset contains the largest number of scenes. Furthermore, unlike other datasets, its dynamic scenes incorporate both radiometric deformation and geometric deformation.

A.6 Positive societal impacts and Limitation

This paper proposes DTGS, a 3D reconstruction method for thermal infrared dynamic scenes, which effectively addresses the challenges posed by the physical characteristics of thermal infrared radiation in dynamic reconstruction. To validate the effectiveness of the proposed method, we construct a thermal infrared dynamic dataset, named Dynamic_LTR, comprising 31 scenes. In the future, with the full open-source release of our code and dataset, the applicability of reconstruction methods to dynamic infrared scenes is expected to be significantly enhanced.

However, DTGS currently achieves suboptimal reconstruction performance in dynamic scenes involving fast camera motion and large rotation, which will be a key focus of our future work.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The core contributions of this work are clearly elaborated in the abstract and introduction sections.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of our work in Appendix A.6.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: In Sections 3.2, 3.3, and 3.4, we provide the full set of assumptions and a complete proof.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We elaborate in detail on the design principles of the DTGS method, and will release all code and data upon paper acceptance.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [N/A]

Justification: We will release all code and data upon paper acceptance.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The training setup is described in detail in Section 4.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: A detailed description is provided in Sections 4.3 and 4.4, as well as in Appendices A.1 through A.4.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: In Section 4.2, we describe the training setup, and in Table 2, we analyze parameters such as training time and rendering speed.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We strictly comply with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss the positive societal impact in Appendix A.6.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All models and datasets mentioned in this paper have been properly cited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We will soon make the models and code presented in this paper publicly available.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.