

Supplementary Materials for “Thermal3D-GS: Physics-induced 3D Gaussians for Thermal Infrared Novel-view Synthesis”

Qian Chen^{1†}, Shihao Shu^{1†}, and Xiangzhi Bai^{1,2,3,*} 

¹ Image Processing Center, Beihang University, Beijing, China

² State Key Laboratory of Virtual Reality Technology and Systems,
Beihang University

³ Advanced Innovation Center for Biomedical Engineering, Beihang University
{19376165, 19375326, jackybxz}@buaa.edu.cn

1 DETAILS OF DATASET

The TI-NSD dataset stands out as the pioneering dataset exclusively designed for thermal infrared novel-view synthesis, encompassing diverse scenarios like indoor, outdoor, and UAV scenes. Specifically, it consists of 7 indoor scenes featuring objects like Scene “Apples” and Scene “Heated” (Heated Bottle), 7 outdoor scenes ranging from Scene “Soccer_Goal” and Scene “Basketball_Court” to Scene “Tall_Building”, and 6 distinct scenes captured by UAV. Fig. 1 illustrates the dataset distribution and the corresponding frame counts for each scene.

2 DETAILS OF RESULTS

2.1 Rendering Efficiency

We present comprehensive Frames Per Second (FPS) testing results in Table 1. The tests were conducted using an NVIDIA RTX 3090 GPU. It is observed that in typical scenarios, Thermal3D-GS achieves high FPS, reaching real-time performance, and notably surpassing real-time in indoor scenes. Additionally, in certain large-scale scenes, including various outdoor and UAV scenarios, our method achieves an average frame rate exceeding 20 FPS, despite the scene’s high complexity.

2.2 Comparison Experiment

Tables 2-6 list the various collected error metrics for our evaluation over all considered techniques. Our method is highlighted in gray, and the optimal indicators are marked in bold.

[†] Co-first authors, contribute equally to this work

* Corresponding author

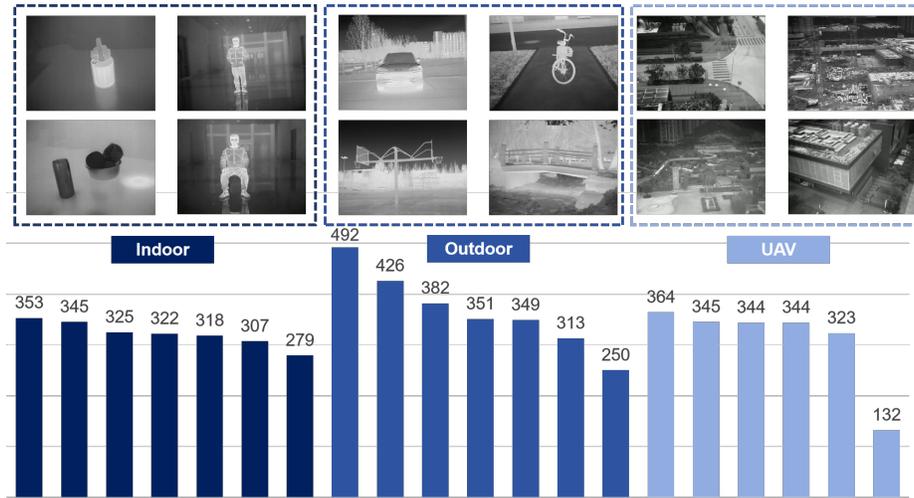


Fig. 1: TI-NSD - the inaugural dataset specifically designed for thermal infrared novel-view synthesis. The line above provides an illustration of three scenario types, with the bottom row indicating the number of frames for each scene.

The results stem from the implementation of InstantNGP [3], encompassing both the base configuration (Base) and a slightly larger network following the authors’ recommendation (Big). Additionally, outcomes from two configurations of 3D-GS [2], executed for 7K and 30K iterations, are included.

Furthermore, we conducted comparative experiments with ThermoNeRF [1], a method for new perspective synthesis using dual modalities of infrared and visible light. ThermoNeRF collects a novel view synthesis dataset named ThermoScenes, comprising 6 indoor scenes and 4 outdoor scenes. However, scenes such as raspberrypi, heater_water_cup, and melting_ice_cup lacked distinct features under the single thermal infrared modality, making it hard for colmap to generate initial point clouds. So they were excluded from the comparison

Table 1: Experiments on FPS.

Indoor		Outdoor		UAV	
Scene	FPS	Scene	FPS	Scene	FPS
Apples	152.91	Chair	21.15	UAV1	23.36
Heated	151.52	Bicycle	29.18	UAV2	23.92
Standing	55.46	Car	16.47	UAV3	19.36
Sitting	53.42	Soccer_Goal	25.19	UAV4	22.60
Corridor	35.19	Basketball_Court	37.51	UAV5	35.25
Wall	98.33	Tall_Building	13.08	UAV6	10.70
Merge	121.95	Bridge	20.16		
Average	95.54	Average	23.25	Average	22.53

Table 2: Scores for Scene Apples, Scene Heated, Scene Standing, Scene Sitting

Scene Method	Apples			Heated			Standing			Sitting		
	SSIM	PSNR	LPIPS									
Plenoxels	0.890	20.11	0.373	0.908	24.73	0.335	0.825	19.82	0.441	0.797	18.23	0.442
INGP-Base	0.931	30.15	0.286	0.938	30.80	0.284	0.920	30.68	0.256	0.933	31.52	0.238
INGP-Big	0.934	30.52	0.278	0.938	30.81	0.283	0.917	30.55	0.267	0.937	31.94	0.243
3D-GS-7k	0.949	30.67	0.292	0.950	30.05	0.282	0.963	34.71	0.235	0.962	33.59	0.229
3D-GS-30k	0.952	32.05	0.284	0.951	30.58	0.279	0.967	36.14	0.219	0.968	36.51	0.211
Ours-7k	0.957	32.32	0.282	0.959	33.17	0.274	0.965	35.06	0.236	0.966	34.92	0.228
Ours-30k	0.961	34.99	0.272	0.963	36.62	0.265	0.968	37.50	0.222	0.970	37.94	0.212

Table 3: Scores for Scene Corridor, Scene Wall, Scene Merge, Scene Chair

Scene Method	Corridor			Wall			Merge			Chair		
	SSIM	PSNR	LPIPS									
Plenoxels	0.898	23.42	0.381	0.909	28.99	0.342	0.843	19.62	0.378	0.732	22.48	0.447
INGP-Base	0.898	19.87	0.361	0.927	23.36	0.270	0.866	22.53	0.340	0.827	27.57	0.294
INGP-Big	0.897	20.38	0.358	0.933	25.52	0.265	0.868	22.52	0.331	0.818	26.79	0.301
3D-GS-7k	0.940	27.53	0.308	0.953	34.12	0.264	0.936	28.28	0.282	0.919	30.12	0.202
3D-GS-30k	0.942	30.63	0.294	0.952	35.28	0.256	0.936	29.70	0.274	0.934	32.32	0.170
Ours-7k	0.940	28.33	0.305	0.960	34.66	0.259	0.950	30.40	0.270	0.921	29.46	0.203
Ours-30k	0.949	32.10	0.290	0.968	39.63	0.244	0.955	33.27	0.257	0.937	33.03	0.172

Table 4: Scores for Scene Bicycle, Scene Car, Scene Soccer_Goal, Scene Basketball_Court

Scene Method	Bicycle			Car			Soccer_Goal			Basketball_Court		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
Plenoxels	0.758	20.99	0.423	0.794	22.98	0.420	0.803	22.92	0.418	0.803	19.53	0.405
INGP-Base	0.854	26.48	0.268	0.868	28.99	0.310	0.850	23.74	0.328	0.874	24.10	0.311
INGP-Big	0.856	26.33	0.260	0.852	27.63	0.319	0.888	29.19	0.265	0.877	24.16	0.308
3D-GS-7k	0.941	30.13	0.195	0.910	27.54	0.249	0.937	30.91	0.222	0.809	17.73	0.346
3D-GS-30k	0.948	31.86	0.177	0.936	31.74	0.201	0.948	33.11	0.194	0.810	19.29	0.346
Ours-7k	0.944	30.70	0.194	0.917	28.03	0.242	0.940	30.96	0.219	0.909	24.21	0.260
Ours-30k	0.954	33.15	0.174	0.942	32.08	0.199	0.952	34.61	0.191	0.945	30.98	0.212

Table 5: Scores for Scene Tall_Building, Scene Bridge, Scene UAV1, Scene UAV2

Scene Method	Tall_Building			Bridge			UAV1			UAV2		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
Plenoxels	0.741	23.28	0.462	0.743	22.84	0.455	0.836	27.16	0.300	0.757	25.16	0.352
INGP-Base	0.751	23.94	0.295	0.832	27.18	0.308	0.587	17.42	0.533	0.633	21.11	0.328
INGP-Big	0.751	23.98	0.290	0.834	27.05	0.304	0.587	17.41	0.529	0.630	21.04	0.311
3D-GS-7k	0.838	21.54	0.278	0.892	26.71	0.257	0.918	29.59	0.175	0.941	30.58	0.117
3D-GS-30k	0.847	24.96	0.273	0.906	28.91	0.225	0.926	32.64	0.154	0.954	33.80	0.098
Ours-7k	0.873	24.19	0.235	0.900	27.66	0.251	0.942	33.11	0.149	0.956	32.91	0.107
Ours-30k	0.944	33.85	0.151	0.918	30.52	0.212	0.956	36.64	0.118	0.966	36.70	0.089

Table 6: Scores for Scene UAV3, Scene UAV4, Scene UAV5, Scene UAV6

Scene Method	UAV3			UAV4			UAV5			UAV6		
	SSIM	PSNR	LPIPS									
Plenoxels	0.622	21.28	0.473	0.831	26.69	0.338	0.839	26.39	0.316	0.794	26.67	0.325
INGP-Base	0.730	22.76	0.325	0.765	22.96	0.368	0.708	19.39	0.412	0.664	21.54	0.455
INGP-Big	0.732	22.78	0.300	0.764	22.95	0.353	0.704	19.42	0.400	0.660	21.33	0.431
3D-GS-7k	0.955	33.48	0.102	0.933	30.10	0.202	0.943	32.30	0.172	0.948	33.11	0.105
3D-GS-30k	0.965	35.69	0.081	0.954	33.87	0.162	0.958	35.33	0.141	0.960	35.74	0.079
Ours-7k	0.955	33.71	0.103	0.951	33.52	0.181	0.951	34.67	0.167	0.955	33.67	0.103
Ours-30k	0.966	36.35	0.082	0.960	36.66	0.159	0.961	36.96	0.142	0.965	37.13	0.081

experiment. We conducted comparative experiments using solely the thermal infrared data and camera pose data provided in “ThermoNeRF”, under identical split conditions. Results were compared with the multimodal method proposed in “ThermoNeRF”, as summarized in Table 7 and Table 8. Visual comparisons are available on Fig. 2

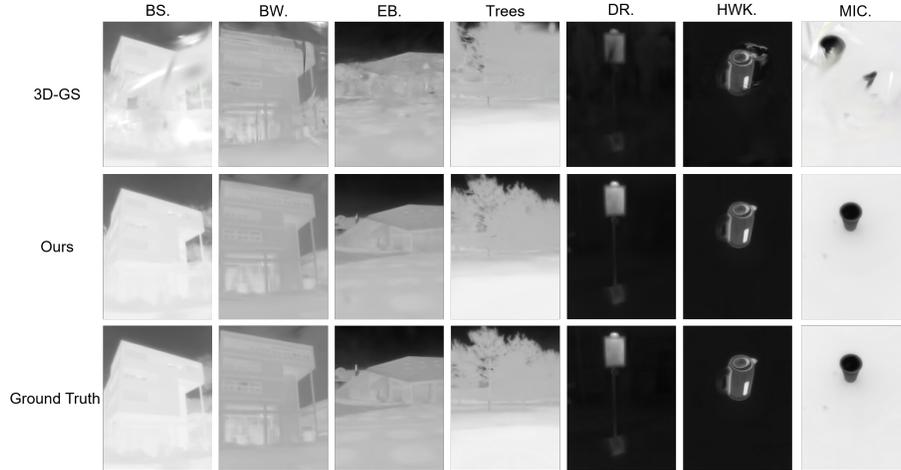


Fig. 2: Visual Comparisons for ThermoScenes Dataset. BS., BW., EB., HWK., MIC. and DR. respectively represent Building (Spring), Building (Winter), Exhibition Building, Heated Water Kettle, Melting Ice Cup and Double Robot. From top to bottom are the results of 3D-GS, ours, and ground truth.

Table 7: PSNR Comparison of **Outdoor** scenes. BS., BW., EB. and Ave. respectively represent Building (Spring), Building (Winter), Exhibition Building and Average.

Scene	Modal	BS.	BW.	EB.	Trees	Ave.
Nerfacto _{th}	th	20.30	22.80	23.88	20.91	21.97
ThermoNeRF	th+rgb	26.63	28.75	33.79	31.07	30.06
3D-GS	th	25.99	34.38	32.05	30.66	30.77
Ours	th	31.55	37.02	36.54	32.17	34.32

2.3 Ablation Study

The comprehensive ablation study results for all 20 scenarios are presented in Table 9-13. Our method is highlighted in gray, and the optimal indicators are marked in bold.

Table 8: PSNR Comparison of **Indoor** scenes. HWK., MIC., DR. and Ave. respectively represent Heated Water Kettle, Melting Ice Cup, Double Robot and Average.

Scene	Modal	HWK.	MIC.	DR.	Ave.
NerfactO _{th}	th	29.25	18.50	10.49	19.41
ThermoNeRF	th+rgb	34.04	32.24	30.75	32.34
3D-GS	th	34.55	33.87	28.42	32.28
Ours	th	35.61	40.99	40.43	39.01

Table 9: Ablation study for Scene Apples, Scene Heated, Scene Standing, Scene Sitting

Scene Method	Apples			Heated			Standing			Sitting		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
3D-GS	0.952	32.05	0.284	0.951	30.58	0.279	0.967	36.14	0.219	0.968	36.51	0.211
3D-GS+ATF	0.959	33.88	0.275	0.954	34.29	0.269	0.968	36.98	0.218	0.970	37.93	0.211
3D-GS+TCM	0.953	32.30	0.283	0.950	32.49	0.273	0.967	36.44	0.220	0.969	36.64	0.211
3D-GS+ \mathcal{L}_{dis}	0.958	33.11	0.274	0.951	32.20	0.972	0.967	36.37	0.220	0.969	36.64	0.212
Ours	0.961	34.99	0.272	0.963	36.62	0.265	0.968	37.50	0.222	0.970	37.94	0.212

Table 10: Ablation study for Scene Corridor, Scene Wall, Scene Merge, Scene Chair

Scene Method	Corridor			Wall			Merge			Chair		
	SSIM	PSNR	LPIPS									
3D-GS	0.942	30.63	0.294	0.952	35.28	0.256	0.936	29.70	0.274	0.934	32.32	0.170
3D-GS+ATF	0.940	31.13	0.299	0.968	39.52	0.243	0.948	32.10	0.264	0.934	32.76	0.171
3D-GS+TCM	0.948	31.65	0.288	0.960	36.29	0.249	0.934	30.20	0.271	0.935	32.42	0.172
3D-GS+ \mathcal{L}_{dis}	0.949	31.44	0.290	0.960	36.15	0.251	0.933	29.87	0.273	0.936	32.52	0.171
Ours	0.949	32.10	0.290	0.968	39.63	0.244	0.955	33.27	0.257	0.937	33.03	0.172

Table 11: Ablation study for Scene Bicycle, Scene Car, Scene Soccer_Goal, Scene Basketball_Court

Scene Method	Bicycle			Car			Soccer_Goal			Basketball_Court		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
3D-GS	0.948	31.86	0.177	0.936	31.74	0.201	0.948	33.113	0.194	0.810	19.293	0.346
3D-GS+ATF	0.952	32.91	0.173	0.939	31.97	0.200	0.950	34.116	0.193	0.925	28.004	0.245
3D-GS+TCM	0.950	31.92	0.174	0.944	32.29	0.195	0.952	34.388	0.187	0.833	20.107	0.339
3D-GS+ \mathcal{L}_{dis}	0.952	32.19	0.175	0.941	32.02	0.202	0.952	34.292	0.189	0.824	20.295	0.327
Ours	0.954	33.15	0.174	0.942	32.08	0.199	0.952	34.609	0.191	0.945	30.981	0.212

Table 12: Ablation study for Scene Tall_Building, Scene Bridge, Scene UAV1, Scene UAV2

Scene Method	Tall_Building			Bridge			UAV1			UAV2		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
3D-GS	0.847	24.96	0.273	0.906	28.91	0.225	0.926	32.64	0.154	0.954	33.80	0.098
3D-GS+ATF	0.925	30.64	0.180	0.914	30.34	0.215	0.955	36.53	0.117	0.966	36.40	0.088
3D-GS+TCM	0.887	27.28	0.233	0.909	29.36	0.221	0.932	33.35	0.144	0.957	34.11	0.098
3D-GS+ \mathcal{L}_{dis}	0.904	28.28	0.202	0.917	29.77	0.210	0.935	33.35	0.142	0.961	34.73	0.091
Ours	0.944	33.85	0.151	0.918	30.52	0.212	0.956	36.64	0.118	0.966	36.70	0.089

Table 13: Ablation study for Scene UAV3, Scene UAV4, Scene UAV5, Scene UAV6

Scene Method	UAV3			UAV4			UAV5			UAV6		
	SSIM	PSNR	LPIPS									
3D-GS	0.965	35.69	0.081	0.954	33.87	0.162	0.958	35.33	0.141	0.960	35.74	0.079
3D-GS+ATF	0.966	36.35	0.081	0.960	36.64	0.156	0.961	36.94	0.138	0.964	37.06	0.079
3D-GS+TCM	0.965	35.84	0.082	0.956	34.78	0.160	0.958	35.97	0.142	0.963	36.17	0.081
3D-GS+ \mathcal{L}_{dis}	0.965	35.71	0.082	0.956	34.30	0.161	0.959	35.57	0.142	0.963	36.41	0.079
Ours	0.966	36.35	0.082	0.960	36.66	0.159	0.961	36.96	0.142	0.965	37.13	0.081

To delve deeper into the influence of the internal composition of ATF and TCM, we conducted additional ablation experiments. Regarding ATF, we investigated the influence of the number of MLP layers, experimenting with configurations of 4, 8, and 16 layers. Average results demonstrate that the 8-layer setup, as utilized in this paper, is optimal, yielding PSNR improvements of 0.396dB and 0.137dB over 4 and 16 layers, respectively. For TCM, we focused on examining the impact of convolutional layer numbers, exploring configurations of 2, 3, and 4 layers. The average PSNR results from the 3-layer convolutional setup used in this paper surpass those of the 2-layer and 4-layer setups by 0.162dB and 0.335dB. Results are presented in Table 14.

Table 14: Ablation study for ATF and TCM.

Scene Method	Indoor			Outdoor			UAV			Average		
	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS
3dgs	0.953	32.98	0.259	0.904	28.89	0.227	0.953	34.51	0.119	0.936	32.01	0.206
3dgs+TCM(2)	0.955	33.46	0.260	0.915	29.55	0.218	0.955	34.95	0.118	0.941	32.54	0.203
3dgs+TCM(3)	0.955	33.72	0.257	0.916	29.68	0.217	0.955	35.04	0.118	0.941	32.70	0.201
3dgs+TCM(4)	0.955	33.44	0.256	0.908	29.17	0.222	0.952	34.84	0.120	0.938	32.37	0.203
3dgs+ATF(4)	0.958	35.04	0.254	0.928	30.56	0.204	0.961	36.56	0.112	0.949	33.93	0.194
3dgs+ATF(8)	0.958	35.12	0.254	0.934	31.53	0.197	0.962	36.65	0.110	0.951	34.33	0.191
3dgs+ATF(16)	0.939	35.03	0.180	0.932	31.23	0.200	0.962	36.65	0.101	0.943	34.19	0.163

2.4 Additional Results

To provide a more comprehensive presentation of the results, we have included interpolation video results in the ZIP file. These results consist of 1,000 frames, all interpolated from the test camera pose. The "Ours" folder contains results obtained using Thermal3D-GS, and the "3D-GS" folder contains results generated by the baseline method 3D-GS. All the results are from the 30,000 training. For conciseness, we showcase only three scenes: Scene UAV1, Scene UAV2 and Scene UAV4.

To comprehensively illustrate the effectiveness of our method, we have supplemented the comparison results from the demonstration video with intercepted images of additional scenes, as depicted in Figure 3.

3 DETAILS OF ATF

We learn the influence of atmospheric transmission effects through an MLP: $(\mu_{abs}, \mu_{sca}, d) = \mathcal{F} ATF(\gamma(x), \gamma(t))$. This MLP maps the position and time of 3D Gaussians to parameters related to atmospheric transmission. Initially, our MLP $\mathcal{F} ATF$ utilizes 8 fully connected layers with ReLU activation functions and 256-dimensional hidden layers to process the input $(\gamma(x), \gamma(t))$. Subsequently, a 256-dimensional feature vector is generated, which passes through three fully connected layers to output μ_{abs}, μ_{sca}, d , completing the learning of atmospheric transmission effects.

References

1. Hassan, M., Forest, F., Fink, O., Miele, M.: Thermonerf: Multimodal neural radiance fields for thermal novel view synthesis (2024)
2. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* **42**(4) (2023)
3. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics* **41**(4), 1–15 (2022)



Fig. 3: We show comparisons of ours to 3D-GS and the corresponding ground truth images in some other scenes. The scenes are, from the top down: Siting, Bicycle, UAV6, UAV5, Wall, Bridge.