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

Supplementary Material

S1. More Examples on Thermal Image Characteristics

S1.1. Radiometric Inconsistency Examples

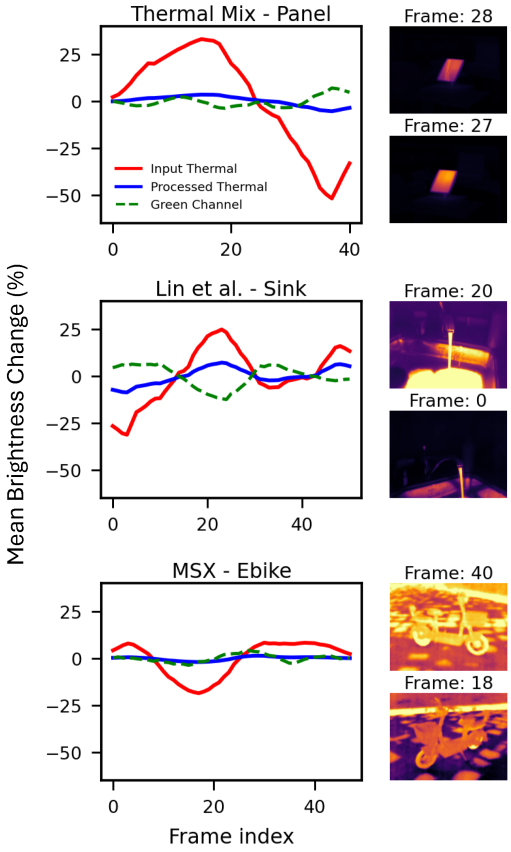

Figure S1. **Radiometric inconsistency across thermal datasets.** Thermal sequences often exhibit frame-to-frame fluctuations, generally influenced by ISP operations, as well as slower radiometric drift that can arise as the sensor gradually warms during capture. Both effects may appear even in static scenes and produce unstable brightness trajectories, as illustrated across the three datasets. These inconsistencies pose a significant challenge for NeRF and 3D Gaussian Splatting methods, which assume photometric consistency across views; variations in thermal intensity can weaken multi-view matching and lead to distorted geometry. In examples from Lin et al. and MSX, we observe global intensity shifts affecting nearly all pixel values, while the ThermalMix example shows a localized change in the foreground object. Our algorithm provides a simple yet effective way to suppress these trends and stabilize temporal behavior.

S1.2. Frequency Content Examples

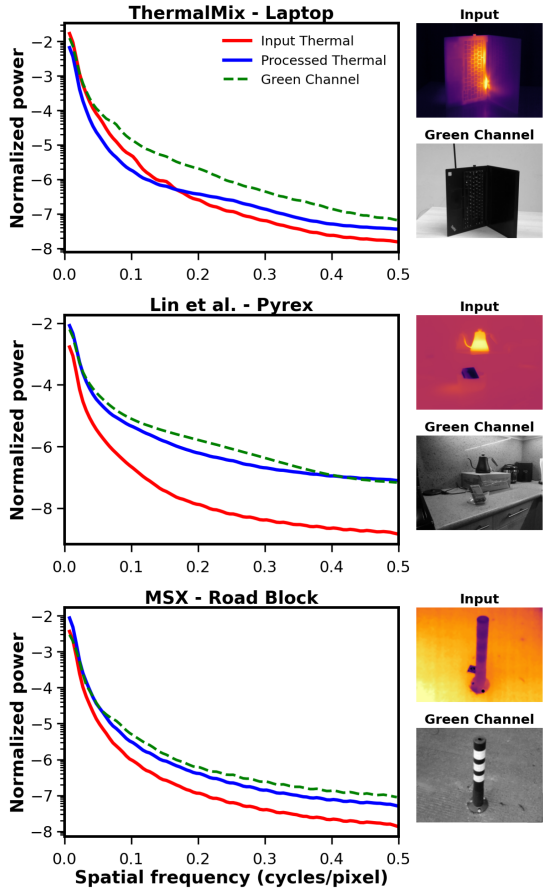

Figure S2. **Spatial frequency characteristics of thermal versus RGB images.** Thermal images generally exhibit weaker mid- and high-frequency components than RGB images. This is partly due to sensor-resolution limits, but also because heat diffuses across surfaces and through the surrounding air, producing the naturally smoother appearance typical of thermal scenes. As a result, fine texture and sharp edges are often diminished, reducing the high-frequency cues that NeRF and 3D Gaussian Splatting rely on for accurate geometry and appearance estimation. Our pipeline increases contrast and strengthens spatial gradients, making features more distinguishable while unavoidably amplifying noise and discretization artifacts. Even so, the enhanced gradients lead to more reliable COLMAP initializations (see Fig. 3) and provide stronger supervision for 3DGS, improving overall reconstruction quality.

S1.3. Histogram Examples

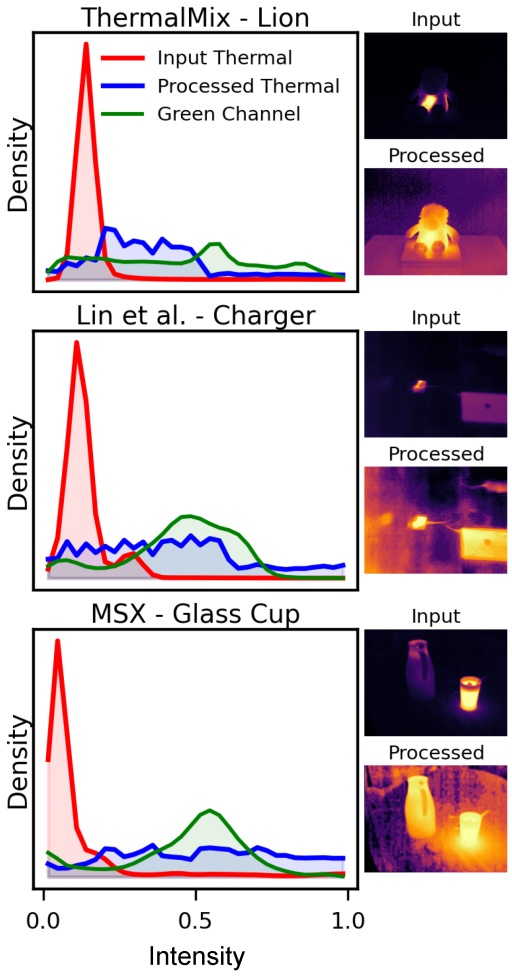

Figure S3. **Intensity distributions of thermal and RGB images.** Raw thermal images typically exhibit highly skewed histograms, where the majority of values accumulate near the lowest intensities. This narrow distribution limits contrast and reduces the visibility of meaningful variations in surface temperature, making it harder for NeRF and 3D Gaussian Splatting to extract useful spatial cues. Our pipeline expands the dynamic range and yields more spread-out intensity distributions, allowing important regions to stand out more clearly. Although this process can make noise more noticeable, the resulting images provide stronger supervision signals and improve reconstruction stability in 3DGS. Example input and processed frames are shown on the right.

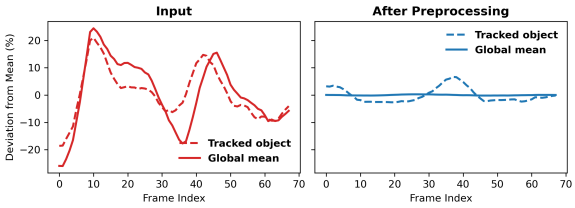

Figure S4. Revisit of Figure 3. Tracked object brightness and global mean before (left) and after (right) preprocessing; local fluctuations follow the global mean and are jointly attenuated.

S2. Local Effects of Preprocessing

To verify that our preprocessing stabilizes thermal intensity locally, we revisit example in Fig. 3. In Fig. S4, we manually track the Apple logo on the laptop across the frames and show how its intensity changes. While raw frames exhibit noticeable frame-to-frame fluctuations, our preprocessing significantly reduces this variability, producing a stable intensity trajectory for the tracked region.

S3. Ablation Study Extended

As shown in Tab. S1, each component of our system addresses a distinct failure mode of thermal reconstruction. Photometric stabilization alone yields small but consistent gains by suppressing frame-to-frame brightness drift, whereas applying contrast enhancement in isolation is unreliable: it helps low-contrast scenes such as Lion and Sink but degrades scenes like Human0 and Ebike where gradients are already strong and contrast boosts amplify noise. When stabilization and enhancement are combined, the effects become complementary, producing a more balanced improvement across datasets. The emission MLP provides the most substantial jump in performance by modeling residual radiometric transients that fixed Gaussian colors cannot capture, raising average PSNR from 23.01 to 24.93 dB. Our full system achieves the best overall results (26.14 dB / 0.88 SSIM), with especially strong improvements on Human0, Ebike, Lion, and Seq.1 while the Sitting scene shows smaller gains due to already high image quality in the baseline reconstructions. Additionally, in Tab. S1, we compare our preprocessing algorithm to traditional histogram equalization, and independently evaluate the effect of each preprocessing step.

S4. Hyperparameters and Weights

Pre-trained weights are provided in our GitHub repository, which also contains the full training code:

https://github.com/NUBIVlab/wild_thermal

Table S1. **Ablation study.** We demonstrate that our preprocessing steps (contrast enhancement as an improvement over traditional histogram equalization, in combination with photometric stabilization) improve 3DGS performance, but not to the level of our method. We then demonstrate the effect of per-frame embedding ablation, with and without pre-processing.

Method	MSX- Ebike		T. Mix - Lion		MVTV - Human		Lin et al. - Sink		Ye et al. - Seq.1		TINSD - Sitting		Avg.	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3DGS (Baseline)	20.45	0.86	19.25	0.71	21.21	0.81	20.81	0.74	28.24	0.83	29.51	**0.88**	22.25	0.81
3DGS + Hist. Eq.	12.54	0.48	20.29	0.79	16.85	0.68	21.02	0.74	28.62	0.81	21.12	0.82	18.86	0.72
3DGS + Contrast Enh.	14.79	0.57	23.97	0.81	17.33	0.68	21.89	0.71	29.90	0.82	20.59	0.80	19.71	0.73
3DGS + Photo. Stab.	23.72	0.87	19.24	0.71	21.65	0.81	21.82	0.76	28.44	0.82	24.50	0.86	22.19	0.81
3DGS + Stab + Enh	22.79	0.86	24.11	**0.82**	22.01	0.84	23.71	0.78	29.90	0.85	22.42	0.85	23.01	0.83
3DGS + Emission MLP	25.73	0.89	24.17	**0.82**	24.22	0.89	**24.57**	0.83	32.12	0.89	25.98	0.87	24.93	0.87
Ours	**25.97**	**0.92**	**24.25**	0.81	**26.18**	**0.90**	24.27	**0.88**	**33.32**	**0.90**	**30.01**	0.87	**26.14**	**0.88**

S5. Video Results

We include full video results in the supplementary materials and show zoomed-in insets for closer inspection. Since Thermal3D-GS is the current state of the art in thermal NVS, these insets focus on comparisons with that method. Note that because these synthesized views are not, generally, part of the test set, there is no ground truth, so large frames are our reconstructions.

To generate novel video frames, we move a virtual camera along pre-determined paths that traverse a subset of the training views. For methods that optimize camera poses, this sometimes produces minor deviations from the intended path, especially in scenes with very limited texture.

S5.1. MVTV - Human0

Figure S4

Figure S5

S5.2. Lin et al. - Sink

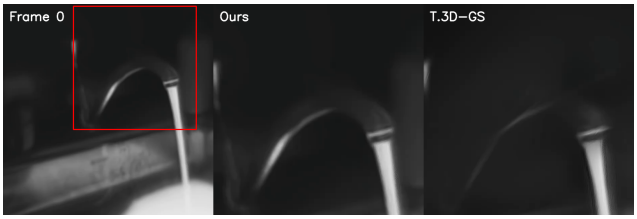

Figure S6

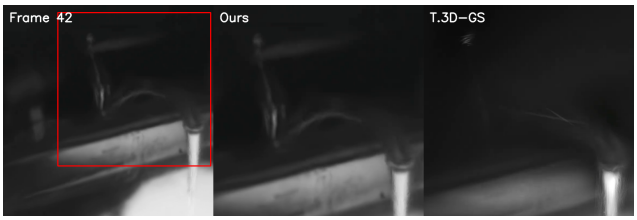

Figure S7

S5.3. MVTV - Mason

Figure S8. Floaters

S5.4. MSX - Ebike

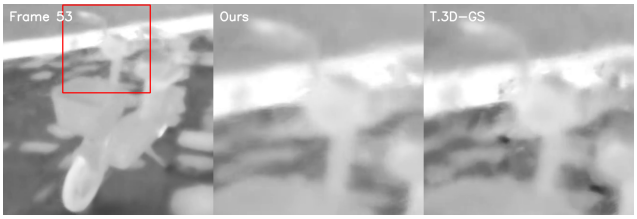

Figure S9

Figure S10

S5.5. ThermalMix - Face

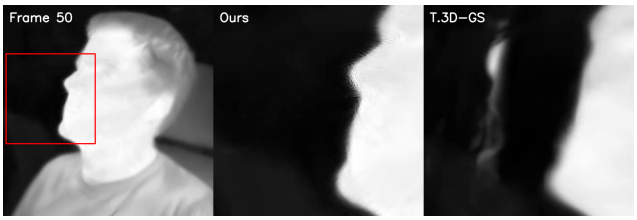

Figure S11

S5.6. MVTV - Chair

Figure S12

S5.7. MSX - Building

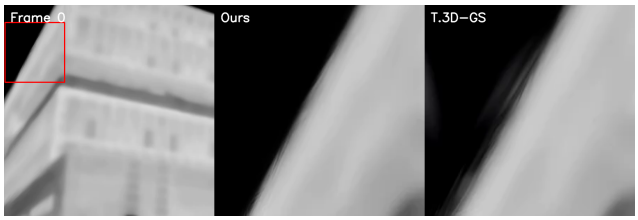

Figure S13

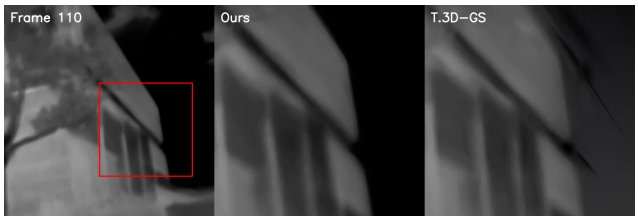

Figure S14

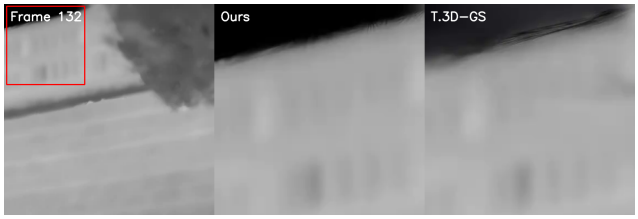

Figure S15

S5.8. MSX - Truck

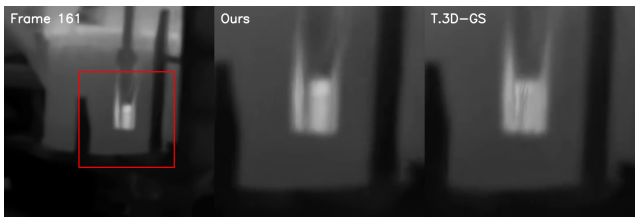

Figure S16

S5.9. Lin et al. - Engine

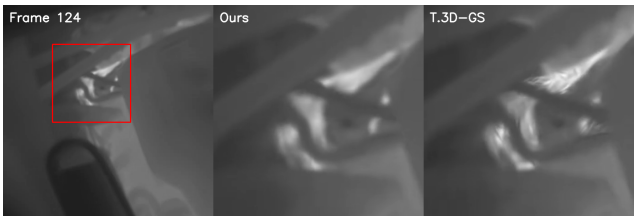

Figure S17

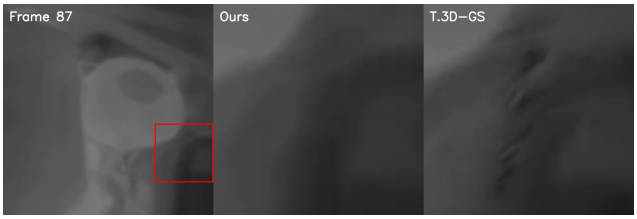

Figure S18

S5.10. MVTV - Parking

Figure S19

S5.11. ThermalMix - Lion

In this video, we crop the reconstructed volume around the lion using a small cubic region to better examine its structure. Without restricting the viewport, the NeRF results contain a large number of floaters. After cropping the volume, it becomes clear that there is no actual density corresponding to the lion's geometry—only view-dependent floaters that explain the training images. This highlights the tendency of NeRF-based methods to overfit thermal inputs and fail to recover the actual geometry.

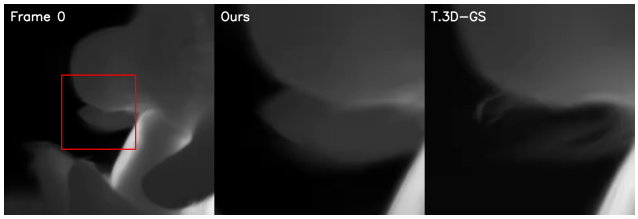

Figure S20

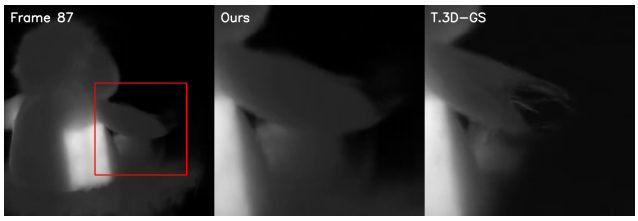

Figure S21

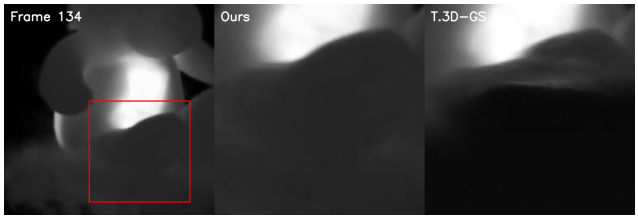

Figure S22