

ThermoNeRF: Multimodal Neural Radiance Fields for Thermal Novel View Synthesis

Mariam Hassan¹, Florent Forest¹, Olga Fink¹, and Malcolm Mielle²

¹ École Polytechnique Fédérale de Lausanne (EPFL)

`firstname.lastname@epfl.ch`

² Schindler AG

`firstname.lastname@schindler.com`

Abstract. Thermal scene reconstruction exhibits great potential for diverse applications, including building energy consumption analysis and non-destructive testing. However, existing methods typically require dense scene measurements and often rely on RGB images for 3D geometry reconstruction, with thermal information being projected post-reconstruction leading to disparities between the geometry and temperatures of the reconstruction and reality. To address this challenge, we propose ThermoNeRF, a novel multimodal approach based on Neural Radiance Fields, capable of rendering new RGB and thermal views of a scene jointly. To overcome the lack of texture in thermal images, we use paired RGB and thermal images to learn scene density, while distinct networks estimate color and temperature information. Furthermore, we introduce ThermoScenes, the first dataset to palliate the lack of available RGB+thermal datasets for scene reconstruction. Experimental results validate that ThermoNeRF shows an average mean absolute error of 1.5°C for temperature estimation, an improvement of over 50% compared to using concatenated RGB+thermal data. `Code` and `dataset` are available through anonymous links.

Keywords: Thermal Imaging · Neural Radiance Fields · 3D Reconstruction · Multimodality

1 Introduction

Since their introduction by Mildenhall et al. [16], Neural Radiance Fields (NeRFs) have achieved great success in 3D reconstruction as well as novel view synthesis and have been extended to diverse tasks [5, 21, 25, 13, 22, 10]. While most NeRF models learn an implicit representation of a scene from a sparse set of RGB images, NeRFs have also been successfully extended to other sensor modalities [26, 19, 23, 26, 6]. However, while it has been demonstrated that NeRFs can learn from more than a single modality at a time (e.g. Poggi et al. [19]), multimodal NeRFs also often depend on the presence of similar features between the modalities [19].

In this paper, we focus on RGB and thermal images. Thermal cameras capture radiation emitted by objects as function of temperature. Since thermal radiation is scattered and reflected from various points, thermal images naturally lack texture and show low edge contrast—an effect known as ghosting [1].

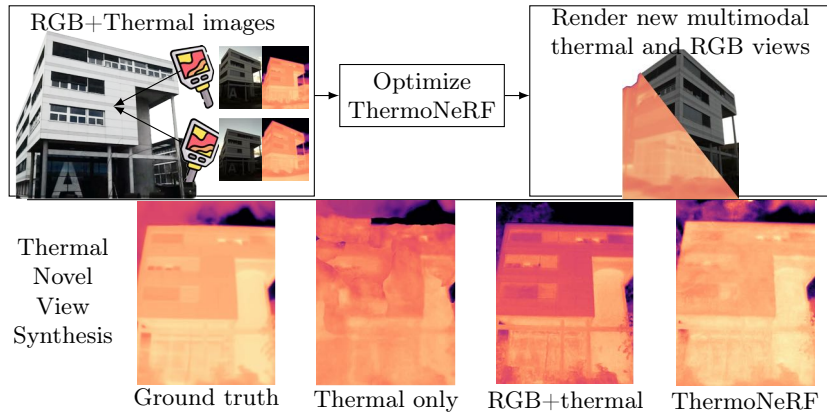


Fig. 1: We propose **ThermoNeRF**, a multimodal NeRF-based approach using paired thermal and RGB images. ThermoNeRF demonstrates enhanced thermal information estimation compared to non-multimodal methods.

Hence, despite the benefits of 3D thermal representations in infrastructure inspection and monitoring [11, 9], non-destructive testing [9], and agriculture [18], models are primarily constructed using photogrammetry, which requires large amounts of RGB data or advanced equipment [15, 7, 14]. Moreover, due to their challenging nature, thermal images are often not considered during geometric reconstruction, and temperature information is only projected post-reconstruction [20], leading to discrepancies between the geometry and temperatures. Additionally, the development of thermal scene reconstruction algorithms is hindered by the lack of benchmark datasets specifically designed for this task.

To address the gaps outlined above, we propose ThermoNeRF (**Thermographic NeRF**), the first multimodal NeRF model capable of rendering unseen views in both RGB and thermal modalities (see Fig. 1). The design of our approach is guided by the specific properties of thermal images. ThermoNeRF uses a shared density MLP that leverages the visual features of the RGB modality to learn the geometry of the scene while ensuring consistency with the thermal measurements. On the other hand, color and thermal information are learned through decoupled MLPs, preventing the influence of colors on the estimated temperatures, and vice-versa. In addition, we also introduce ThermoScenes, the first paired thermal-RGB dataset comprising ten diverse scenes—six indoor and four outdoor scenes.

Our contributions are summarized as follows:

- We propose ThermoNeRF, the first multimodal NeRF capable of rendering both thermal and RGB views jointly.
- We conduct extensive experiments to validate how ThermoNeRF’s architecture allows the RGB modality to guide the density estimation for thermal reconstruction while preventing information leakage between both modalities.

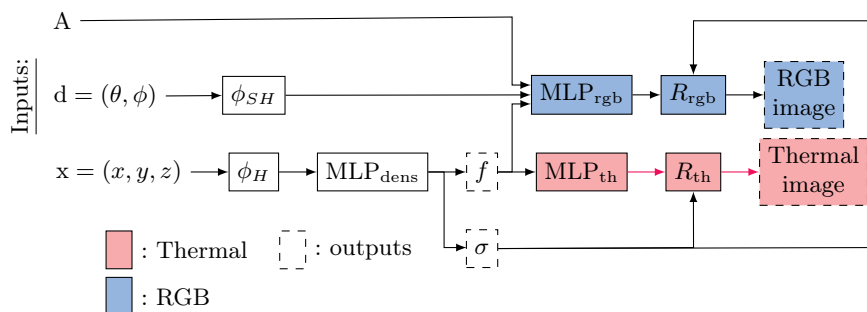


Fig. 2: ThermoNerf architecture; in red are parts of the network related to the generation of thermal images, while blue represents the parts related to the generation of RGB images. The MLP_{th} is only dependent on the intermediate features f as input. A is appearance embeddings that accounts for difference in exposure for RGB images.

- We provide ThermoScenes, the first RGB+thermal images dataset for 3D scene reconstruction and novel view synthesis, featuring ten scenes with diverse temperature ranges and types of objects.
- Finally, we present a comprehensive evaluation covering both temperature estimation and reconstruction quality on unseen poses.

Our results demonstrate improved temperature estimation with no loss in reconstruction fidelity when compared to models trained using only thermal images or concatenated RGB+thermal images as input.

2 ThermoNeRF

In this section, we present ThermoNeRF, a NeRF model capable of learning an implicit scene representation in both thermal and RGB, from a sparse set of RGB and thermal images. To synthesize realistic and accurate RGB and thermal novel views, ThermoNeRF retains geometric features captured by the RGB modality, while ensuring that temperature estimates are independent of scene color variations. Our design choices are guided by two properties of thermal images: 1) Thermal images are inherently *soft and textureless*. 2) Temperatures are *independent of the viewing direction*.

Refer to Fig. 2 for a flowchart of the method; implementation of the model and evaluations can be found online¹.

2.1 Thermal Image Rendering

As depicted in Fig. 3, while RGB information can be view-dependent due to non-Lambertian effects, thermal cameras measure thermal radiation and convert it to temperature, which is invariant with respect to the viewing direction. Therefore,

¹ <https://github.com/Schindler-EPFL-Lab/thermo-nerf>

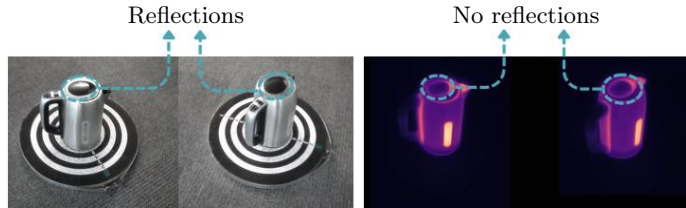


Fig. 3: Non-Lambertian effects—i.e. light reflections—present in the RGB images (left) depend on the angle of view and are not present in thermal images (right). Furthermore, textures and edge features in the thermal images are soft due to the ghosting effect, as opposed to the sharpness of the RGB image and its background.

ThermoNeRF uses separate MLPs for each modality: MLP_{th} predicts temperatures along each ray, while MLP_{rgb} predicts RGB values. Since temperatures are independent of the viewing direction, MLP_{th} only takes the intermediate features f of MLP_{dens} as input, while MLP_{th} receives f , ϕ_{SH} and A —see Fig. 2.

Furthermore, as seen in the thermal images in Fig. 3, the ghosting effect makes thermal images soft and textureless. Consequently, while textures and sharp geometric features in RGB images allow NeRF models to accurately estimate scene density, NeRF struggles to learn densities solely from thermal images—see Fig. 4b for experimental results of training Nerfacto on thermal images only. Therefore, in ThermoNeRF, a single density MLP_{dens} is shared by the modalities, enabling the network to construct a geometric representation informed by RGB information while ensuring geometric consistency with the thermal modality.

Formally, ThermoNeRF is defined as follows:

$$\begin{aligned} [\sigma, f] &= \text{MLP}_{\text{dens}}(\phi_H(\mathbf{x})), \\ c &= \text{MLP}_{\text{rgb}}(f, \phi_{SH}(d), A), \\ t &= \text{MLP}_{\text{th}}(f), \end{aligned} \tag{1}$$

where t denotes the predicted temperature values and A is appearance embedding.

2.2 Loss Functions

We use separate reconstruction loss functions for each modality. The final reconstruction loss \mathcal{L} is expressed as the sum of the Mean Square Error (MSE) losses of the RGB and thermal outputs, respectively denoted \mathcal{L}_{rgb} and \mathcal{L}_{th} . Similar to Nerfacto, we add the interlevel and distortion losses, which were initially used in MIP-NeRF 360 [2], to optimize the proposal sampler and to reduce distortions respectively. Therefore, our final loss is defined as follows:













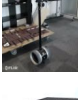
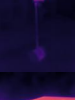






$$\mathcal{L} = \mathcal{L}_{\text{rgb}} + \mathcal{L}_{\text{th}} + \mathcal{L}_{\text{dist}} + \mathcal{L}_{\text{interl}}. \tag{2}$$

3 ThermoScenes Dataset

We introduce ThermoScenes, a new dataset comprising paired RGB and thermal images for ten scenes (four outdoor and six indoor) with diverse temperature ranges. Images are collected using a FLIR One Pro LT [8] which captures calibrated pairs of RGB and thermal images. The FLIR One Pro LT thermal camera operates within $-20\text{ }^{\circ}\text{C}$ to $120\text{ }^{\circ}\text{C}$, with a thermal accuracy of $\pm 3\text{ }^{\circ}\text{C}$.

We assess the precision to be approximately $\pm 0.14\text{ }^{\circ}\text{C}$. To calculate the precision of our thermal camera, we fix the camera at a position and capture 20 consecutive images of a RaspberryPi connected to power for at least 10 minutes to ensure thermal stability. Our precision is then the average standard deviation of the temperatures of each pixel across the 20 images.

Table 1: Summary of the collected ThermoScenes dataset showing example views with paired RGB and thermal images and the temperature range for each scene.

Scene	RGB	Thermal	#views	Temp. range	Scene	RGB	Thermal	#views	Temp. range
Heated Water Cup			95 (train) 13 (test)	23.8°C 68.6°C	Hot Water Kettle			77 (train) 10 (test)	20.0°C 87.3°C
Frozen Ice Cup			133 (train) 19 (test)	-16.2°C 23.1°C	Melting Ice Cup			85 (train) 12 (test)	0.4°C 25.5°C
Building (Spring)			107 (train) 15 (test)	-62.5°C* 19.7°C	Building (Winter)			84 (train) 12 (test)	-15.7°C 15.6°C
Double Robot			83 (train) 11 (test)	21.0°C 29.3°C	Raspberry Pi			111 (train) 15 (test)	22.3°C 41.8°C
Exhibition Building			119 (train) 16 (test)	-11.3°C 14.0°C	Trees			73 (train) 10 (test)	-8.4°C 11.8°C

*The values are outside of its operating range on the clear sky. Due to the sky’s variations of atmospheric radiation, the readings are subject to systematic inaccuracies [12]. We clip these values to the minimum of the operating range ($-20\text{ }^{\circ}\text{C}$). However, our data can be post-processed for more accurate estimates on the sky [12, 24].

Tab. 1 provides a summary of the ThermoScenes dataset. The test set for each scene comprises one-eighth of the total number of images, with test views uniformly sampled across the camera poses. ThermoScenes is publicly available².

² <https://zenodo.org/records/10835108>

4 Experiments

We conduct a comprehensive evaluation of rendered novel views using ThermoScenes. This section details the evaluation metrics, compares baselines, and discusses results. Refer to the supplementary material for results on RGB.

4.1 Evaluation Metrics

We assess the accuracy of the rendered temperatures by calculating the Mean Absolute Error (MAE). Given that thermal images often display uniform temperature distributions outside the regions of interest (ROI), calculating the MAE across the entire image may disproportionately emphasize the ambient background temperature. From an application standpoint, the temperature of the ROI is crucial for temperature assessment [4]. Therefore, we also report the MAE_{roi} , the MAE computed over the region of interest. We use Otsu’s method [17] to determine the optimal threshold that distinguishes the ROI pixels from the background. We also report the standard image quality metrics to evaluate the reconstruction quality for RGB and thermal images.

4.2 Baselines

We conduct experiments to evaluate the two design choices, i.e. (1) the necessity of multimodality, as thermal images alone cannot accurately estimate the scene density due to lack of texture details, and (2) the decoupling of the RGB and thermal modalities, stemming from the physical independence between temperature and color information. Given that ThermoNeRF is based on Nerfacto, we define two baseline methods derived from Nerfacto: $\text{Nerfacto}_{\text{th}}$ and $\text{Nerfacto}_{\text{rgb+th}}$.

Nerfacto_{th} is trained with thermal inputs only. It processes thermal images as single-channel grayscale images through Nerfacto’s standard pipeline. It highlights the importance of incorporating the RGB modality.

Nerfacto_{rgb+th} takes both RGB and thermal modalities as inputs by concatenating them into four-channel images and optimizes the concatenated RGB-thermal images without employing separate MLPs for optimizing each modality. This highlights the importance of disjoint optimization of the modalities.

4.3 Thermal View Synthesis

Tab. 2 provides a comparative analysis of our method’s performance for novel view synthesis for the thermal modality on the ThermoScenes dataset.

Tab. 2 reports the average per-pixel MAE_{roi} and MAE of thermal views synthesis on the test set. With an average MAE of 0.66°C , ThermoNeRF significantly outperforms $\text{Nerfacto}_{\text{rgb+th}}$ (1.54°C) and $\text{Nerfacto}_{\text{th}}$ (2.89°C). When trained on thermal images alone, $\text{Nerfacto}_{\text{th}}$ fails to learn an accurate representation of the scene (as visible in Fig. 4) and has the lowest image quality and temperature estimation results. On the other hand, thanks to the concatenation of RGB

Table 2: Quantitative comparison of our method (ThermoNeRF) versus Nerfacto_{th} and Nerfacto_{rgb+th} on thermal novel view synthesis across the ten scenes in ThermoScenes.

Metric	Method	Heated		Freezing		Melting		Building		Double Raspberry		Exhibition		Trees	Avg
		Water	Cup	Water	Kettle	Ice	Cup	Ice	Cup	(Spring)	(Winter)	Robot	Pi		
PSNR \uparrow	Nerfacto _{th}	23.68	29.25	23.34	18.50	20.30	22.80	10.49	18.08	23.88	20.91	21.12			
	Nerfacto _{rgb+th}	29.76	31.80	22.9	32.70	20.60	28.47	29.82	24.3	27.74	31.48	27.96			
	ThermoNeRF	32.05	34.04	30.67	32.24	26.63	28.75	30.75	31.80	33.79	31.07	31.18			
SSIM \uparrow	Nerfacto _{th}	0.71	0.89	0.95	0.93	0.91	0.87	0.45	0.71	0.94	0.92	0.83			
	Nerfacto _{rgb+th}	0.83	0.91	0.96	0.98	0.89	0.89	0.89	0.82	0.95	0.94	0.91			
	ThermoNeRF	0.92	0.94	0.97	0.98	0.92	0.88	0.95	0.96	0.97	0.94	0.94			
MAE _{roi} \downarrow	Nerfacto _{th}	13.57	5.18	6.75	12.27	6.36	1.80	2.85	4.82	1.29	1.60	5.65			
	Nerfacto _{rgb+th}	5.35	3.25	10.33	1.26	6.54	0.86	1.06	1.62	1.00	0.31	3.16			
	ThermoNeRF	2.10	2.76	3.26	1.57	1.88	0.66	0.91	1.28	0.31	0.25	1.50			
MAE \downarrow	Nerfacto _{th}	1.82	4.19	1.67	1.97	6.74	1.87	5.33	2.25	1.56	1.49	2.89			
	Nerfacto _{rgb+th}	0.87	1.54	2.01	0.46	6.59	0.89	0.55	1.15	0.97	0.34	1.54			
	ThermoNeRF	0.53	0.71	0.57	0.29	2.40	0.76	0.34	0.27	0.35	0.33	0.66			

and thermal information, Nerfacto_{rgb+th} obtains better results than Nerfacto_{th}: MAE_{roi} of 3.16°C, against 5.65°C for Nerfacto_{th}. However, the estimated temperatures are impacted by the joint optimization of both modalities in the same MLP, as evidenced in Fig. 5 where the error in temperature prediction is high on the ROI. On the contrary, the average MAE_{roi} and MAE for ThermoNeRF are 1.5°C and 0.66°C, respectively, marking a significant improvement over the second-best result of 3.16°C and 1.54°C.

Additionally, we visually compare examples of test views rendered by each of the baseline methods as well as by ThermoNeRF (see Fig. 4). We observe noisy renderings with Nerfacto_{th}, while Nerfacto_{rgb+th} generates sensible reconstruction of the geometry—some scenes such as the Double Robot scene being noisier than others. However, temperature predictions are influenced by the RGB modality, leading to biased temperature values and sharper edges compared to the ground truth for the Building (Spring). Moreover, we can see in the rendered thermal images, elements from the RGB view that are invisible in the ground truth thermal images—this is especially visible in the Raspberry Pi scene with the mouse and keyboard (Fig. 4c). This indicates that the RGB channels have influenced the temperature estimates, as both modalities are optimized within the same MLP. Similar renderings for all ten datasets are shown in the supplementary material.

In Fig. 5, we visualize the per-pixel absolute errors in temperature prediction and observe highest errors for all models at the edges. This is due to the ghosting effect available in thermal images where the edges are blurry. We also can observe that for Building Spring (top-left), the sky has more errors than the Exhibition Building scene (top-right). That is due to having a mix of clear sky and cloudy images on the Building Spring scene, as opposed to just a cloudy sky in the Exhibition Building scene. The inconsistencies of thermal radiation measured from the sky leading to less accurate thermal predictions. Overall, ThermoNeRF exhibits the best thermal prediction results while Nerfacto_{th} shows the worst thermal predictions. Nerfacto_{rgb+th} shows higher errors across the scenes than ThermoNeRF due to the contamination of the thermal information by the RGB.

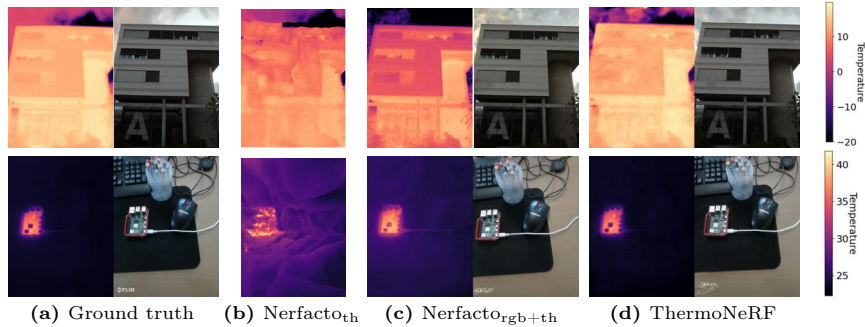


Fig. 4: Comparison of examples of thermal and RGB renderings of unseen poses for the scenes Building Spring, and Raspberry Pi. Note that $\text{Nerfacto}_{\text{th}}$ has no RGB output. ThermoNeRF is closest to ground-truth thermal images, while preserving RGB quality.

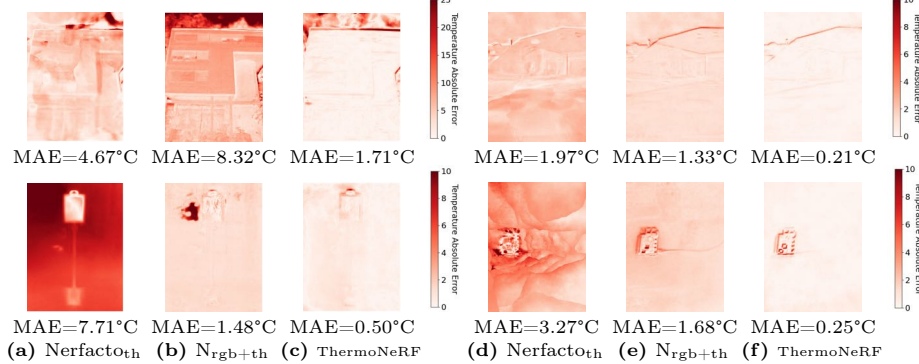


Fig. 5: Per-pixel absolute errors in temperature estimation for renderings of unseen poses for two outdoor scenes: (top) Building Spring and Exhibition Building and two indoor scenes: (bottom) Double Robot and RaspberryPi. We observe fewer errors on ThermoNeRF than the baselines. Note that $N_{\text{rgb}+\text{th}}$ stands for $\text{Nerfacto}_{\text{rgb}+\text{th}}$.

5 Conclusion

In this work, we propose ThermoNeRF, a novel multimodal approach leveraging NeRF for rendering novel RGB and thermal views of a scene. We curated a new dataset specifically designed for RGB+thermal scene reconstruction and our experimental findings show that ThermoNeRF excels in synthesizing thermal images, achieving an average MAE of 1.5°C, representing more than 50% improvement over a baseline concatenating the RGB and thermal modalities. Moreover, a compelling avenue for future research is exploring the adaptability of ThermoNeRF across different frameworks such as recent anti-aliasing NeRF variant, ZIP-NeRF [3]. Another potential limitation of our method is its dependence on paired RGB and thermal images. Collecting paired RGB-thermal images can pose practical challenges since careful calibration of the cameras is needed to align these modalities accurately. Future work will focus on developing methods able to train ThermoNeRF using unpaired RGB and thermal image data.

References

1. Bao, F., Jape, S., Schramka, A., Wang, J., McGraw, T.E., Jacob, Z.: Why are thermal images blurry, (2023). eprint: 2307.15800 (physics.optics) (cit. on p. 1).
2. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. CVPR (2022) (cit. on p. 4)
3. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. ICCV (2023) (cit. on p. 8)
4. Brooke, C.: Thermal Imaging for the Archaeological Investigation of Historic Buildings. Remote Sensing **10**(9) (2018) (cit. on p. 6). <https://doi.org/10.3390/rs10091401>. <https://www.mdpi.com/2072-4292/10/9/1401>
5. Chang, W., Zhang, Y., Xiong, Z.: Depth Estimation From Indoor Panoramas With Neural Scene Representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023) (cit. on p. 1)
6. Chen, C., Richard, A., Shapovalov, R., Ithapu, V.K., Neverova, N., Grauman, K., Vedaldi, A.: Novel-view acoustic synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023) (cit. on p. 1)
7. De Luis-Ruiz, J.M., Sedano-Cibrián, J., Pérez-Álvarez, R., Pereda-García, R., Salas-Menocal, R.B.: Generation of 3D Thermal Models for the Analysis of Energy Efficiency in Buildings. In: Cavas-Martínez, F., Marín Granados, M.D., Mirálbes Buil, R., de-Cózar-Macías, O.D. (eds.) Advances in Design Engineering III, pp. 741–754. Springer International Publishing, Cham (2023) (cit. on p. 2)
8. FLIR ONE Pro Thermal Imaging Camera for Smartphones | Teledyne FLIR, <https://www.flir.com/products/flir-one-pro/> (cit. on p. 5).
9. Gade, R., Moeslund, T.: Thermal cameras and applications: A survey. Machine Vision and Applications **25**, 245–262 (2014) (cit. on p. 2). <https://doi.org/10.1007/s00138-013-0570-5>
10. Hu, B., Huang, J., Liu, Y., Tai, Y.-W., Tang, C.-K.: NeRF-RPN: A general framework for object detection in NeRFs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 23528–23538 (2023) (cit. on p. 1)
11. Kim, H., Lamichhane, N., Kim, C., Shrestha, R.: Innovations in Building Diagnostics and Condition Monitoring: A Comprehensive Review of Infrared Thermography Applications. Buildings **13**(11) (2023) (cit. on p. 2). <https://doi.org/10.3390/buildings13112829>
12. Kruczek, T.: Conditions for use of long-wave infrared camera to measure the temperature of the sky. Energy **283**, 128466 (2023) (cit. on p. 5). <https://doi.org/10.1016/j.energy.2023.128466>. <https://www.sciencedirect.com/science/article/pii/S0360544223018601>
13. Liu, F., Zhang, C., Zheng, Y., Duan, Y.: Semantic Ray: Learning a Generalizable Semantic Field with Cross-Projection Attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023) (cit. on p. 1)
14. Martin, M., Chong, A., Biljecki, F., Miller, C.: Infrared thermography in the built environment: A multi-scale review. Renewable and Sustainable Energy Reviews **165**, 112540 (2022) (cit. on p. 2). <https://doi.org/10.1016/j.rser.2022.112540>
15. Maset, E., Fusiello, A., Crosilla, F., Toldo, R., Zorretto, D.: PHOTOGAMMETRIC 3D BUILDING RECONSTRUCTION FROM THERMAL IMAGES. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **IV-2-W3**, 25–32 (2017) (cit. on p. 2). <https://doi.org/10.5194/isprs-annals-IV-2-W3-25-2017>. <https://isprs-annals.copernicus.org/articles/IV-2-W3/25/2017/> (visited on 02/26/2024)

16. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In: ECCV (2020) (cit. on p. 1)
17. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics* **9**(1), 62–66 (1979) (cit. on p. 6). <https://doi.org/10.1109/TSMC.1979.4310076>
18. Parihar, G., Saha, S., Giri, L.L.: Application of infrared thermography for irrigation scheduling of horticulture plants. *Smart Agricultural Technology* **1**, 100021 (2021) (cit. on p. 2). <https://doi.org/10.1016/j.atech.2021.100021>. <https://www.sciencedirect.com/science/article/pii/S2772375521000216>
19. Poggi, M., Ramirez, P.Z., Tosi, F., Salti, S., Mattoccia, S., Stefano, L.D.: Cross-Spectral Neural Radiance Fields. In: 2022 International Conference on 3D Vision (3DV), pp. 606–616 (2022) (cit. on p. 1). <https://doi.org/10.1109/3DV57658.2022.00071>. <https://ieeexplore.ieee.org/abstract/document/10044414> (visited on 02/14/2024)
20. Ramón, A., Adán, A., Javier Castilla, F.: Thermal point clouds of buildings: A review. *Energy and Buildings* **274**, 112425 (2022) (cit. on p. 2). <https://doi.org/10.1016/j.enbuild.2022.112425>. <https://www.sciencedirect.com/science/article/pii/S0378778822005965>
21. Tosi, F., Tonioni, A., De Gregorio, D., Poggi, M.: NeRF-Supervised Deep Stereo. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 855–866 (2023) (cit. on p. 1)
22. Xu, C., Wu, B., Hou, J., Tsai, S., Li, R., Wang, J., Zhan, W., He, Z., Vajda, P., Keutzer, K., Tomizuka, M.: NeRF-Det: Learning Geometry-Aware Volumetric Representation for Multi-View 3D Object Detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 23320–23330 (2023) (cit. on p. 1)
23. Zhang, J., Zhang, F., Kuang, S., Zhang, L.: NeRF-LiDAR: Generating Realistic LiDAR Point Clouds with Neural Radiance Fields. In: AAAI Conference on Artificial Intelligence (AAAI) (2024) (cit. on p. 1)
24. Zhao, C., Zhang, L., Zhang, Y.: All-sky longwave radiation modelling based on infrared images and machine learning. *Building and Environment* **238**, 110369 (2023) (cit. on p. 5). <https://doi.org/10.1016/j.buildenv.2023.110369>
25. Zhi, S., Laidlow, T., Leutenegger, S., Davison, A.J.: In-Place Scene Labelling and Understanding with Implicit Scene Representation. In: ICCV (2021) (cit. on p. 1)
26. Zhu, H., Sun, Y., Liu, C., Xia, L., Luo, J., Qiao, N., Nevatia, R., Kuo, C.-H.: Multimodal Neural Radiance Field. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), pp. 9393–9399 (2023) (cit. on p. 1). <https://doi.org/10.1109/ICRA48891.2023.10160388>