

Under-Canopy Terrain Reconstruction in Dense Forests Using RGB Imaging and Neural 3D Reconstruction Supplementary Materials

Refael Sheffer¹ Chen Pinchover¹ Haim Zisman² Dror Ozeri¹ Roei Litman
¹Rafael Advanced Defense Systems inc., Israel ²Bar-Ilan University, Israel

This supplementary document provides additional details and extended results complementing the main text. First, Section 1 provides detail on the reconstruction of the different scenes, as well as a result on thermal data. Then, Section 2 provides additional visualization of different loss functions. Next, Section 3 describes in detail the construction of the synthetic scene that was used in the main paper. Finally, Section 4 provides more details on the person detection task.

Table 2 is based on Table 1 from the AOS paper [6], where in addition to summary of the forest type, acquisition date and number of targets, it also combines details for all three sections, and is referenced accordingly.

Note that there are two oblique view 3D animations included with this document: first, the segmentation canopy removal, and second the stem detection. Each animation is based on the scene depicted in its corresponding figure in the main text (segmentation on F5, stem detection on F9), depicting the scene from different angles.

1. Reconstruction results

Our pipeline includes two reconstruction stages, *i.e.* SfM and NeRF training. We discuss the challenges and failure cases of each stage below.

SfM results As mentioned, we use COLMAP [7] to reconstruct the relative camera pose, as well as camera intrinsics. In Table 2 below we report key statistics from the Structure-from-Motion (SfM) stage, including reprojection error and gray level range. As mentioned in the main text, we see that the F7 flight is characterized with very low dynamic range, which is a likely culprit for the failure of COLMAP on it.

NeRF reconstructions NeRF reconstruction was done using the modified version of Instant-NGP [5] that was mentioned in the text. As mentioned, ‘broadleaf’ and ‘mixed’ vegetation were harder to penetrate in the given image capture settings, and some of the people in the AOS

dataset were not reconstructed for these cases. Table 2 includes a column indicating the number of people that are at least partially visible, and were annotated for detection by us.

Table 3 and Table 4 depict the top view rendering of all NeRF reconstructions, both before and after removal of the canopy. In addition, we also include the ‘integral image’ view from the AOS dataset [6], which gives sense of the location of the people, and also the difficulty of each scene from the thermal point of view. As mentioned in the main text, all ‘broadleaf’ and ‘mixed’ scenes show lower quality results, but only in F1, F3 and F4 some people are not visible at all. These visualizations illustrate the effectiveness of our approach in revealing terrain and objects under dense foliage, as well as its limitations in low-light or highly occluded conditions.

Thermal Scene Rendering. Although this paper primarily focuses on RGB data, we also demonstrate the applicability of our approach to thermal imagery, which was the main focus in [6]. Figure 1 depicts one image from the F6 scene, as well as two under canopy imaging results. First, looking at the original thermal image, one can still observe some human subjects, but occlusions remain and may hinder automated SAR person detection. Second, the AOS method [6], which addresses this limitation by design, reveals all persons in a single occlusion-free image. Finally, our method also successfully reveals the people, while additionally reconstructing parts of the surrounding environment, such as trees and shrubs. These results suggest that thermal imagery can be leveraged not only for detecting warm-bodied targets but also for reconstructing broader scene context. A more thorough quantitative comparison of person detection using thermal data is out of the scope of this paper.

2. Loss functions visualizations

Table 5 includes additional visual comparisons of four loss functions: L1, L2, Huber [2], and RAW [4]. As men-

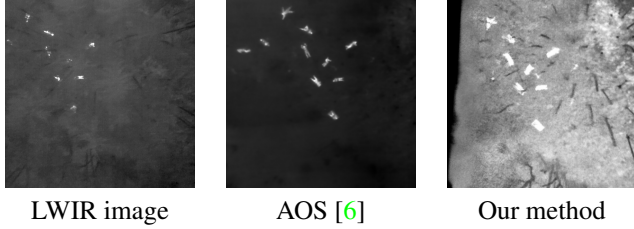


Figure 1. Comparison of the AOS [6] method and our method on the F6 seen from the dataset. See text for more details.

tioned in the main text, L1 is the best off-the-shelf candidate for under-canopy reconstruction, while the proposed RAW loss seems to be more robust, and reconstructs people even in the hardest scenes, such as F2. L2 loss on the other hand, and sometimes also Huber loss, fail to reliably reconstruct even the easiest scenes. These results visually highlight the trade-offs between robustness to dynamic range and preservation of fine details, supporting our choice of RAW loss for challenging illumination scenarios.

3. Synthetic data

Since the data in [6] contains only airborne drone images, we are unable to qualitatively assess the quality of the reconstruction. To address this gap, even partially, we include results on a synthetically generated scene based on renderings of dense forest of 3D tree models using Blender [1].

The virtual scene covers approximately $30m \times 30m$ of forested area. We place 16 trees of the same species, Japanese larch (*Larix kaempferi*), arranged on a regular 4×4 grid with 7 m spacing between neighboring trunks. Each tree has a height of 16 m and a canopy diameter of 6.7 m. In addition, to enrich the vegetation structure, we place three smaller under-canopy trees of the same species, as well as fallen trees. To emulate the presence of man-made objects under the canopy, we place four red-white motorcycles on the forest floor at different locations beneath the trees. Illumination is set to diffuse, non-direct sunlight (overcast-like conditions), consistent with the acquisition conditions recommended in the main paper. See Figure 2 for screen capture of the scene before it was rendered.

The camera is modeled as a pinhole camera with focal length $f = 3.56$ mm and an image resolution of 1882×3566 pixels. We simulate three parallel flight lines of length 17 m each, flown at a constant altitude of 20 m above the ground, resulting in a total of 150 rendered views. The camera is oriented with a slight off-nadir viewing angle along the flight direction, mimicking a typical UAV acquisition pattern.

For evaluation, we render two versions of the scene: First, the full scene containing both canopy and ground

objects, and second, a canopy-free version in which all tree geometry is removed while keeping the terrain and under-canopy objects fixed. The canopy-free rendering serves as a noise-free ground-truth surface image, against which we compare the reconstructions produced by our method and by 3DGS.

4. Person detection

We provide additional details on the person detection pipeline used in our experiments. As mentioned, the design was intentionally simple, relying on pretrained networks and unsupervised components where applicable. Our approach combines object proposals from a pretrained detector, with unsupervised anomaly scoring person detection.

First, a COCO pretrained yolo11n [3] detector processes 800×800 image patches with a 25% stride to produce candidate boxes. This detector was chosen since it empirically yields high recall on the dataset if a relaxed threshold of 0.01 is used. Proposals near patch boundaries or below a 0.01 confidence threshold are discarded using non-maximum suppression. Each retained proposal box is resized to 224×224 crop and embedded using a ResNet-50 network pretrained on ImageNet, yielding a 2048-dimensional feature vector. An MLP autoencoder (whose architecture is detailed in Table 1) attempts to reconstruct this vector, and its mean squared reconstruction error constitutes the anomaly detection statistic (*i.e.* abnormality score).

Part	Layers
Encoder	Linear($input_dim \rightarrow hidden_dim$) \rightarrow ReLU
	Linear($hidden_dim \rightarrow hidden_dim/2$) \rightarrow ReLU
Decoder	Linear($hidden_dim/2 \rightarrow hidden_dim$) \rightarrow ReLU
	Linear($hidden_dim \rightarrow input_dim$)

Table 1. Architecture of the autoencoder anomaly detector. Since $input_dim = 2048$, we use $hidden_dim = 512$.

The autoencoder was trained on ResNet embeddings of YOLO detections crops, taken from images without persons (F8–F11) and the full scene of F0 with the canopy. Notably, this training regime does not rely on any labeled detection data and can be easily adapted to new scenes under the reasonable assumption that no visible persons are present in the autoencoder training set. This training data was split into two sets: ‘null’ set which is used to train the autoencoder weights, and ‘calibration’ used to select the threshold for detection, set as the 95th percentile of reconstruction errors. Figure 3 illustrates the distribution of reconstruction losses for the ‘null’ and ‘calibration’, with the threshold indicated by a dashed vertical line. During inference, boxes whose reconstruction loss exceeds this threshold are labeled as persons.

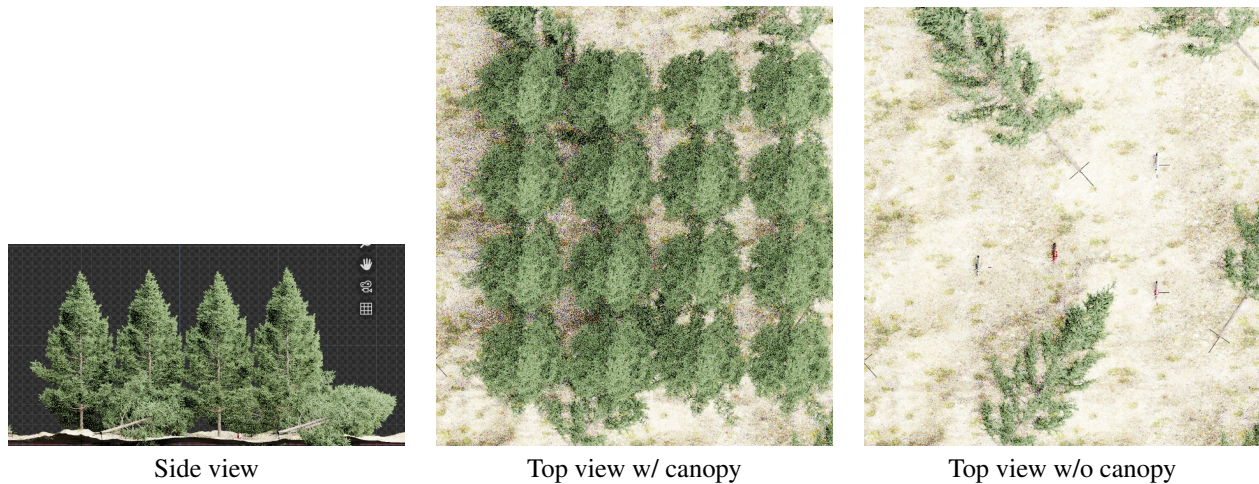


Figure 2. Screenshots of the scene from Blender [1], before rendering. See text for details.

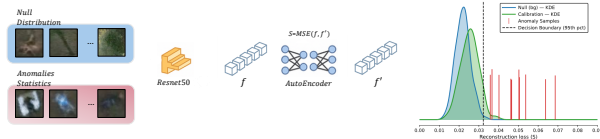


Figure 3. Flow diagram of the person detector, from left to right. See section 4 for details.

Table 2 includes the full person detection statistics across all scenes, where the final AP value is shown in the main text. As mentioned, false negatives (FN) include only persons at least partially visible in the reconstruction, consistent with the AOS evaluation protocol [6]. Performance varies significantly across scenes, with failures concentrated in flights with broadleaf canopies (*i.e.* F3 and F4).

References

- [1] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 2, 3
- [2] Peter J Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution*, pages 492–518. Springer, 1992. 1, 7
- [3] Glenn Jocher and Jing Qiu. Ultralytics yolo11. <https://github.com/ultralytics/ultralytics>, 2024. 2
- [4] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. NeRF in the dark: High dynamic range view synthesis from noisy raw images. *CVPR*, 2022. 1, 7
- [5] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–15. ACM, 2022. 1
- [6] David C. Schedl, Indrajit Kurmi, and Oliver Bimber. Airborne optical sectioning for revealing persons in forested environments using synthetic aperture imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:332–346, 2021. 1, 2, 3, 4, 5
- [7] Johannes L. Schönberger. *Robust Methods for Accurate and Efficient 3D Modeling from Unstructured Imagery*. PhD thesis, ETH Zürich, 2018. 1

ID	Forest	Capture month	Image count	Person count	Reproj. Err. (px)	Gray level			Visible /Total	Person detection		
						Low	High	Range		TP	FP	FN ^d
F0 ^a	conifer	Oct	411	3	1.42	7	183	176	3/3	3	5	1
F1 ^b	broadleaf	Oct	207	10	1.38	33	190	157	4/10	4	0	0
F2	broadleaf	Oct	325	10	1.38	40	185	145	10/10	10	0	0
F3	mixed	Oct	403	6	1.43	11	228	217	5/6	1	5	3
F4	mixed	Oct	378	6	1.45	6	166	160	2/6	0	3	1
F5	conifer	Nov	328	10	1.42	1	148	147	10/10	10	1	0
F6	conifer	Nov	322	10	1.40	2	187	185	10/10	6	0	3
F7 ^c	broadleaf	Nov	31	2	N/A	1	82	81	0/2	N/A	N/A	N/A
F8	broadleaf	Jan	357	0	1.38	6	178	172	0/0	–	1	–
F9	broadleaf	Jan	359	0	1.45	7	205	198	0/0	–	0	–
F10	conifer	Apr	398	0	1.41	6	169	163	0/0	–	2	–
F11	conifer	Apr	417	0	1.42	1	161	160	0/0	–	1	–

Table 2. Consolidated summary of forest plot metadata (from Table 1 in [6]), SfM statistics, NeRF reconstruction visibility counts, and person detection results for all flights. ^a 1 m \times 2 m spacing. ^b Flight aborted early. ^c 5 m circular synthetic aperture; 31 images. SfM failed on this scene, so it was not included in the person detection evaluation. ^d FN counts include only persons at least partially visible in our reconstructions (similar to the AOS protocol [6]).



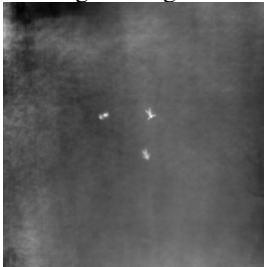


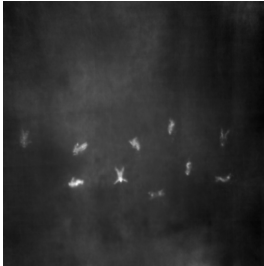


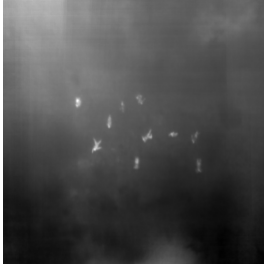

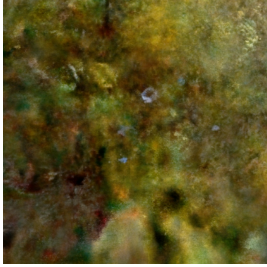
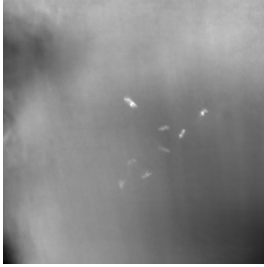


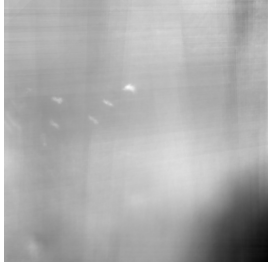
	Full scene	Canopy removed	thermal AOS integral image
F0			
F1			
F2			
F3			
F4			

Table 3. NeRF reconstructions (before and after canopy removal) for "F" scenes, alongside thermal integral images from the AOS dataset [6]. F7 is excluded due to SfM failure; additional results are in Table 4 below.

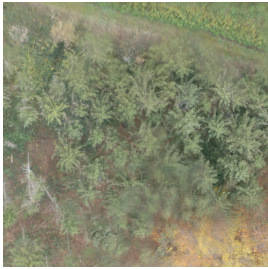



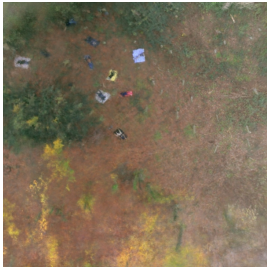



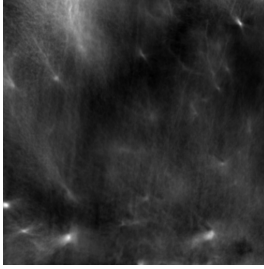

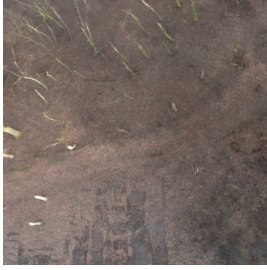
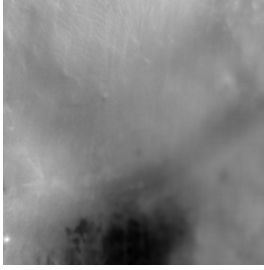


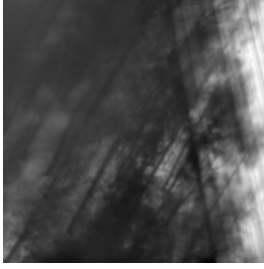


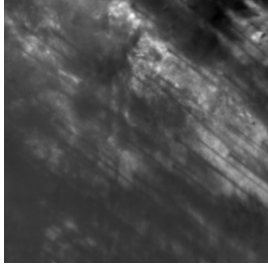
	Full scene	Canopy removed	thermal AOS integral image
F5			
F6			
F8			
F9			
F10			
F11			

Table 4. Additional NeRF reconstruction results for remaining scenes (see Table 3 for context).

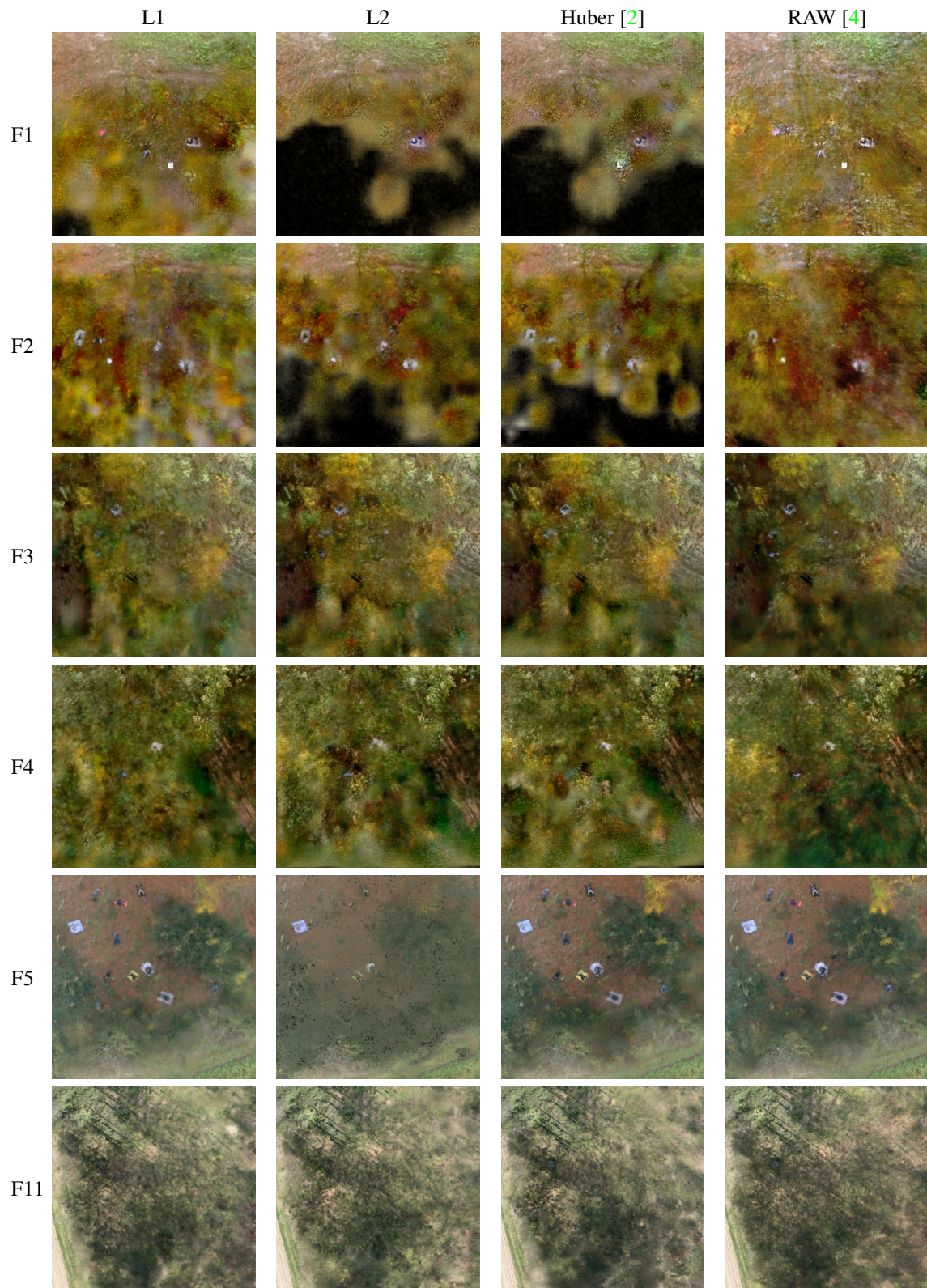


Table 5. Visual comparison of four loss functions: L1, L2, Huber [2] and Low light loss from RawNeRF [4]