SatUnreal: A High-Precision Synthetic Dataset for Satellite Stereo Matching via Unreal Engine

Supplementary Material

A. Extended Qualitative Samples across Diverse Terrains

We present a comprehensive visual gallery of SatUnreal across all eight geographic domains (Fig. S1 through Fig. S8). For each domain, we provide RGB images and corresponding disparity ground truths across three systematic baselines (50m, 129m, and 200m). These samples illustrate our dataset's capacity to provide high-precision supervision while stress-testing specific network vulnerabilities:

- **Urban Environments (Figs. S1, S7 and S8):** High-density models (California, Toronto) feature extreme elevation changes and severe occlusions, which are precisely masked by our Two-Step LineTrace algorithm. Venice adds complex canal structures to ensure broad urban transferability.
- **Low-Texture Regions (Figs. S4 and S5):** Deserts and fields introduce severe low-texture ambiguities and repetitive patterns, directly challenging photometric consistency and global context extraction.
- **Complex Natural Terrains (Figs. S2, S3 and S6):** Canyons and Forests test the recovery of fine-grained details across large disparity ranges, while Coastal areas evaluate geometric matching robustness at heterogeneous land-water interfaces.

B. Cross-Dataset Generalization: Qualitative Analysis

To visually substantiate our zero-shot transferability claims, Figs. S9 to S11 compare Selective-IGEV predictions trained exclusively on SatUnreal, US3D, and WHU-Stereo across all three test domains.

Despite significant domain shifts—from the nadir-focused real-world imagery of WHU-Stereo to the highly off-nadir synthetic urban canyons of SatUnreal—all models exhibit remarkable cross-domain compatibility, reasonably retaining the target geometric structures.

Critically, the model trained purely on our synthetic SatUnreal dataset yields reconstructions visually on-par with those trained on large-scale real-world datasets. In highly occluded and boundary regions, it often maintains clear structural definitions, effectively mitigating ambiguous predictions. This strongly validates that SatUnreal's physically accurate rendering and flawless occlusion masks successfully bridge the Sim-to-Real gap, matching the generalization power of real-world benchmarks.

B.1. Cross-Dataset Generalization: DLNR Architecture

We extend this qualitative evaluation to the DLNR architecture (Figs. S12 to S14) to explore specific model-data dynamics:

- **Robustness of SatUnreal:** Consistent with the Selective-IGEV results, the SatUnreal-trained DLNR generalizes strongly. It successfully recovers smooth and coherent terrain structures on the SatUnreal and WHU-Stereo test sets, demonstrating highly competitive performance compared to the in-domain baseline, particularly in preserving sharp topological transitions.
- **Architecture-Data Mismatch on US3D:** Interestingly, all DLNR models exhibit sub-optimal qualitative performance on the US3D test set. We attribute this to a fundamental mismatch: DLNR is inherently designed to produce globally smooth disparity maps via cascaded refinement, which contrasts with the intricate high-frequency variations often found in US3D's LiDAR-derived ground truths. This mismatch results in overly smoothed, "blob-like" predictions.

While DLNR excels at minimizing global EPE by preventing catastrophic outliers (as shown in Tab. 2), it inherently struggles to reconstruct the noisy details of unrefined real-world LiDAR datasets. Nevertheless, our synthetic SatUnreal dataset remains the most effective supervisory source for maximizing structural coherency across diverse domains.

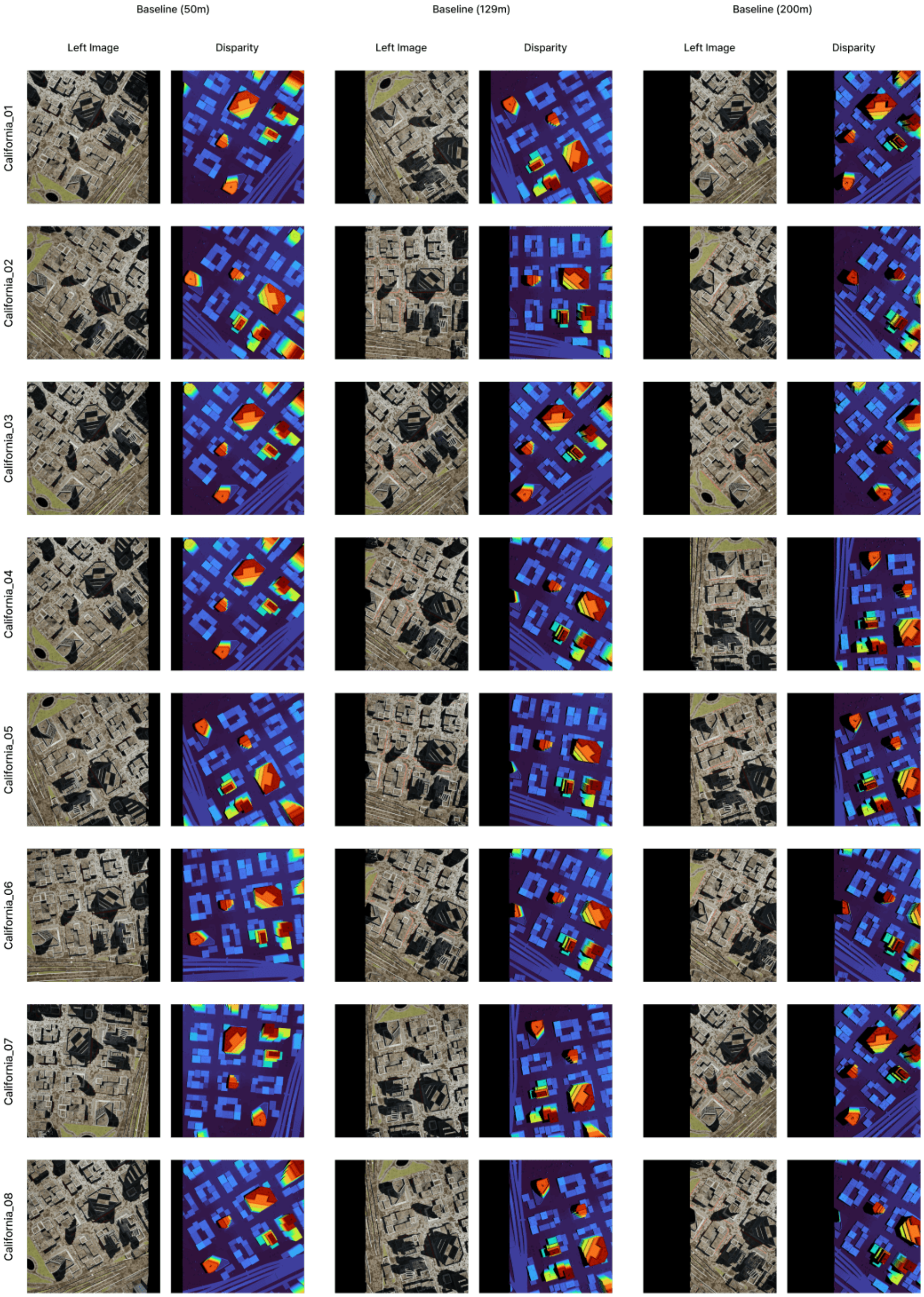

Figure S1. Qualitative samples for the California (Urban) domain.

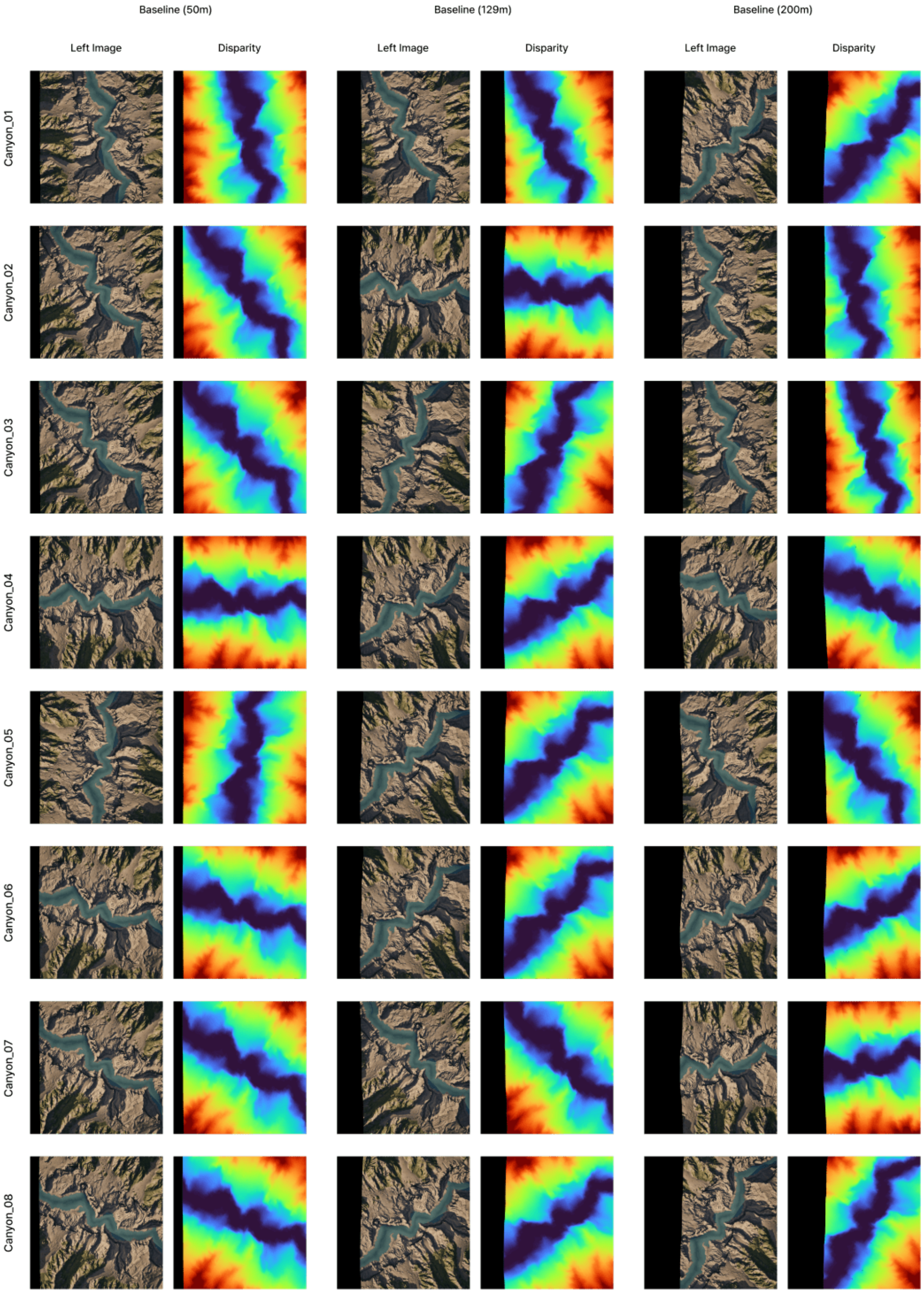

Figure S2. Qualitative samples for the Canyon (Natural) domain.

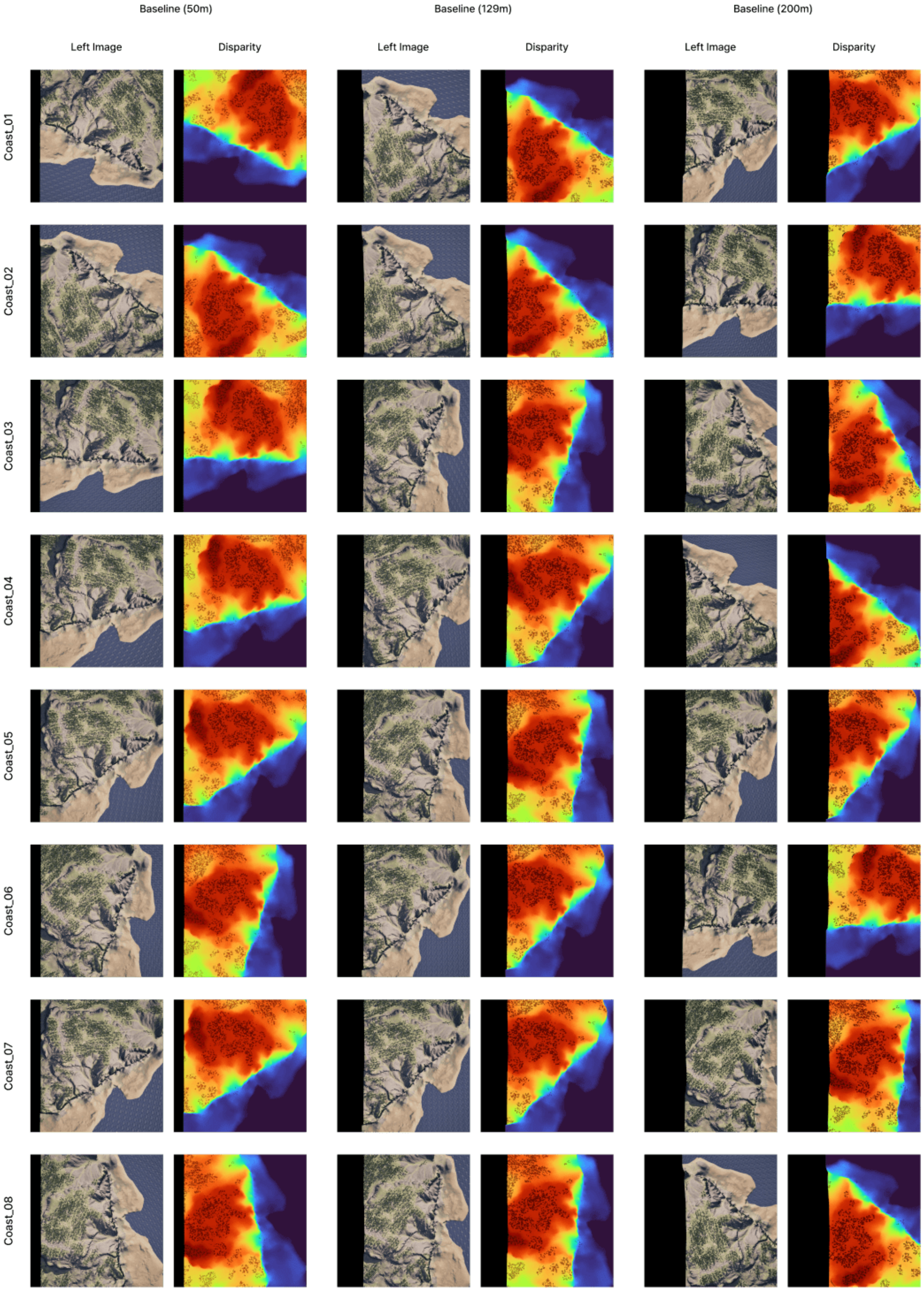

Figure S3. Qualitative samples for the Coast (Natural) domain.

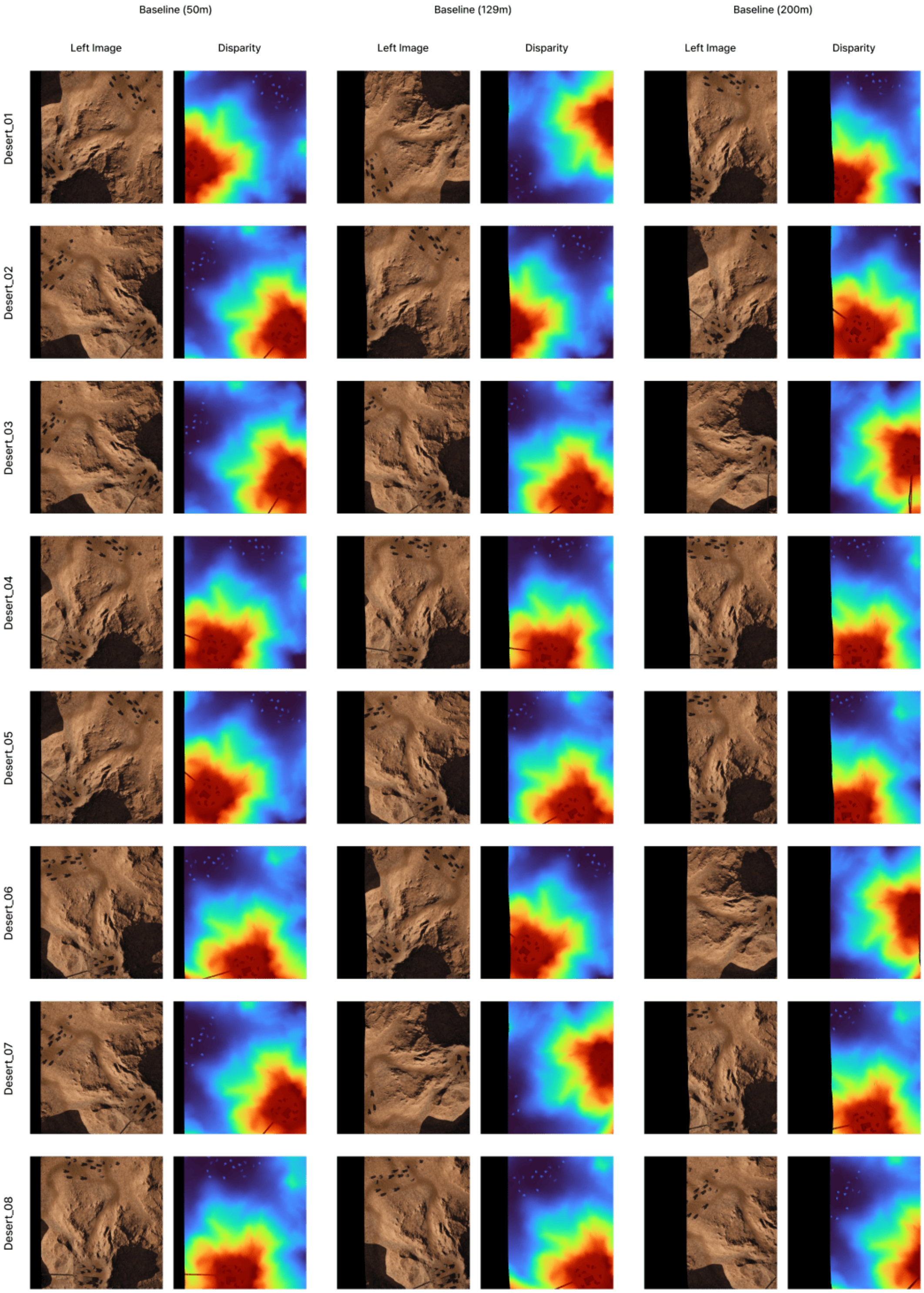

Figure S4. Qualitative samples for the Desert (Natural) domain.

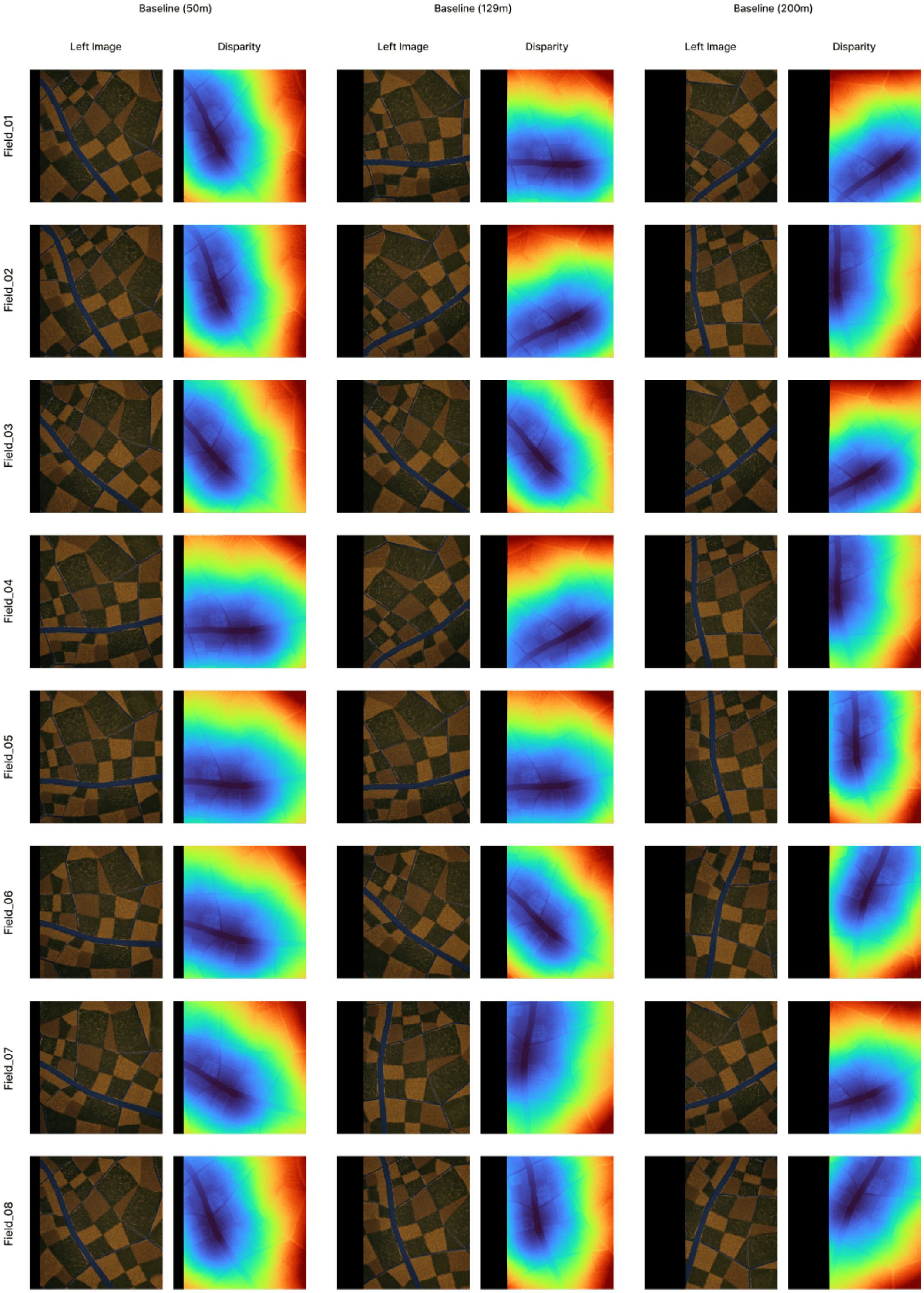

Figure S5. Qualitative samples for the Field (Natural) domain.

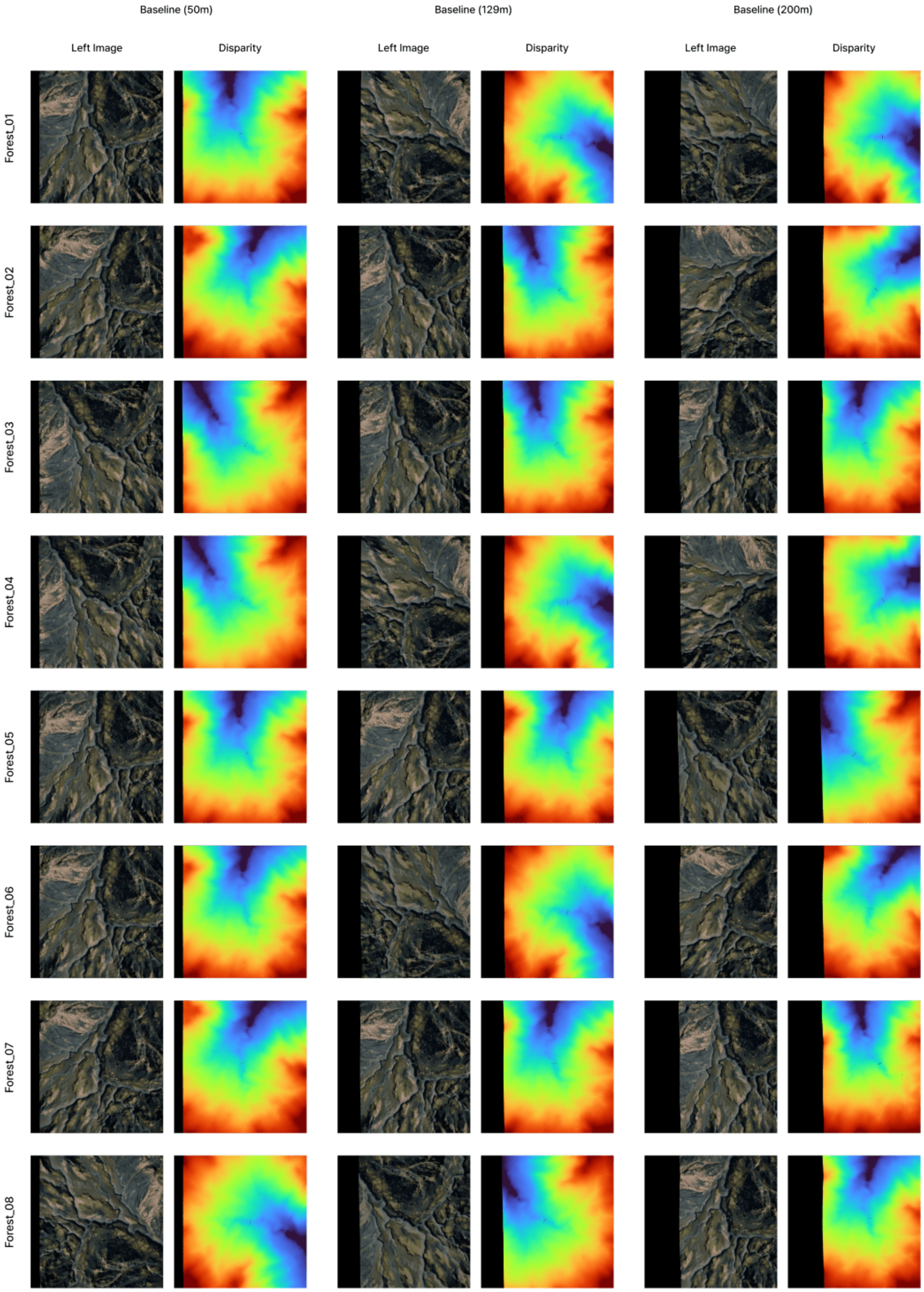

Figure S6. ualitative samples for the Forest (Natural) domain.

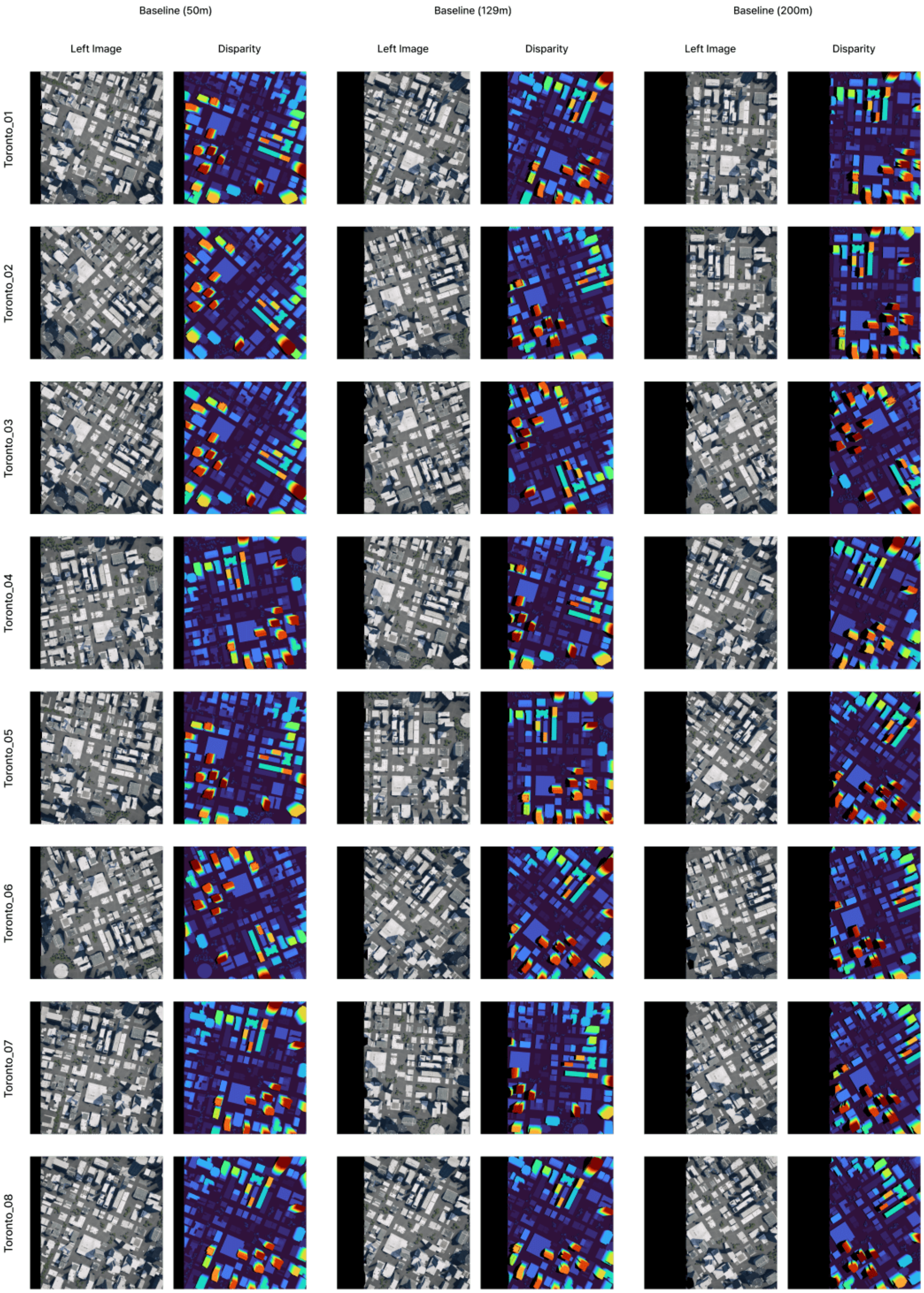

Figure S7. Qualitative samples for the Toronto (Urban) domain.

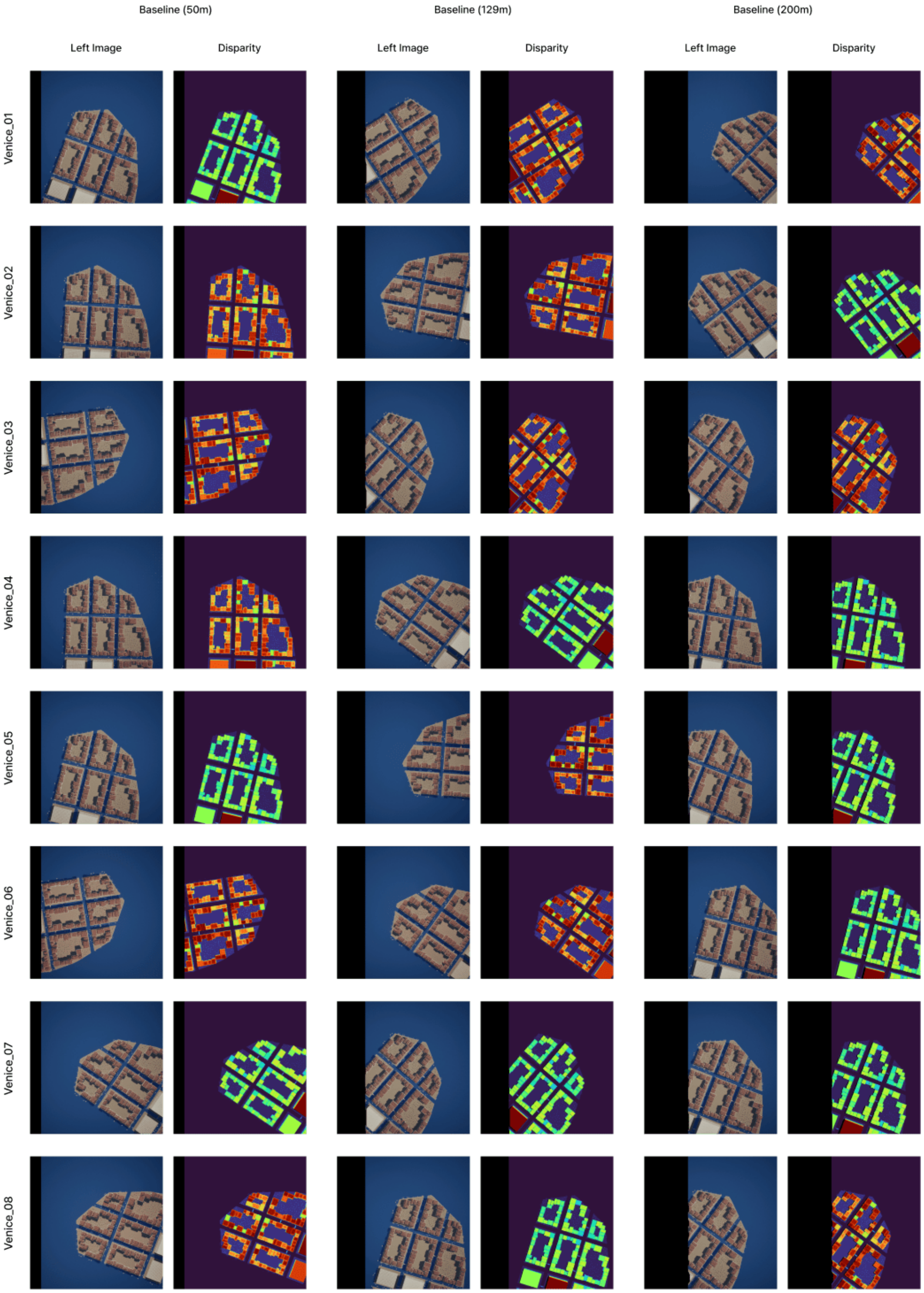

Figure S8. Qualitative samples for the Venice (Urban) domain.

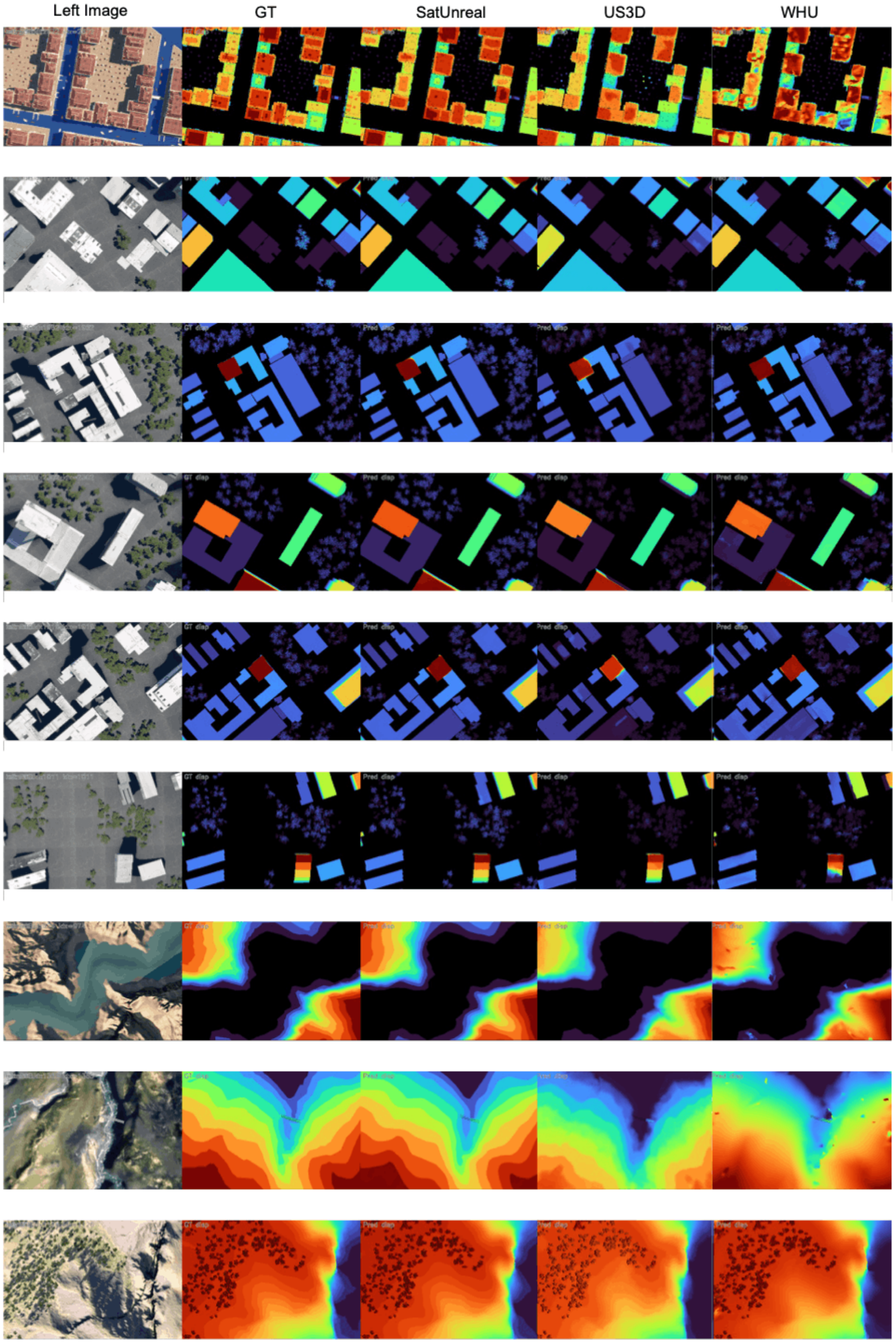

Figure S9. Qualitative evaluation on the SatUnreal test set.From left to right: Left image, Ground Truth (GT), and inference results from models trained on SatUnreal, US3D, and WHU-Stereo.

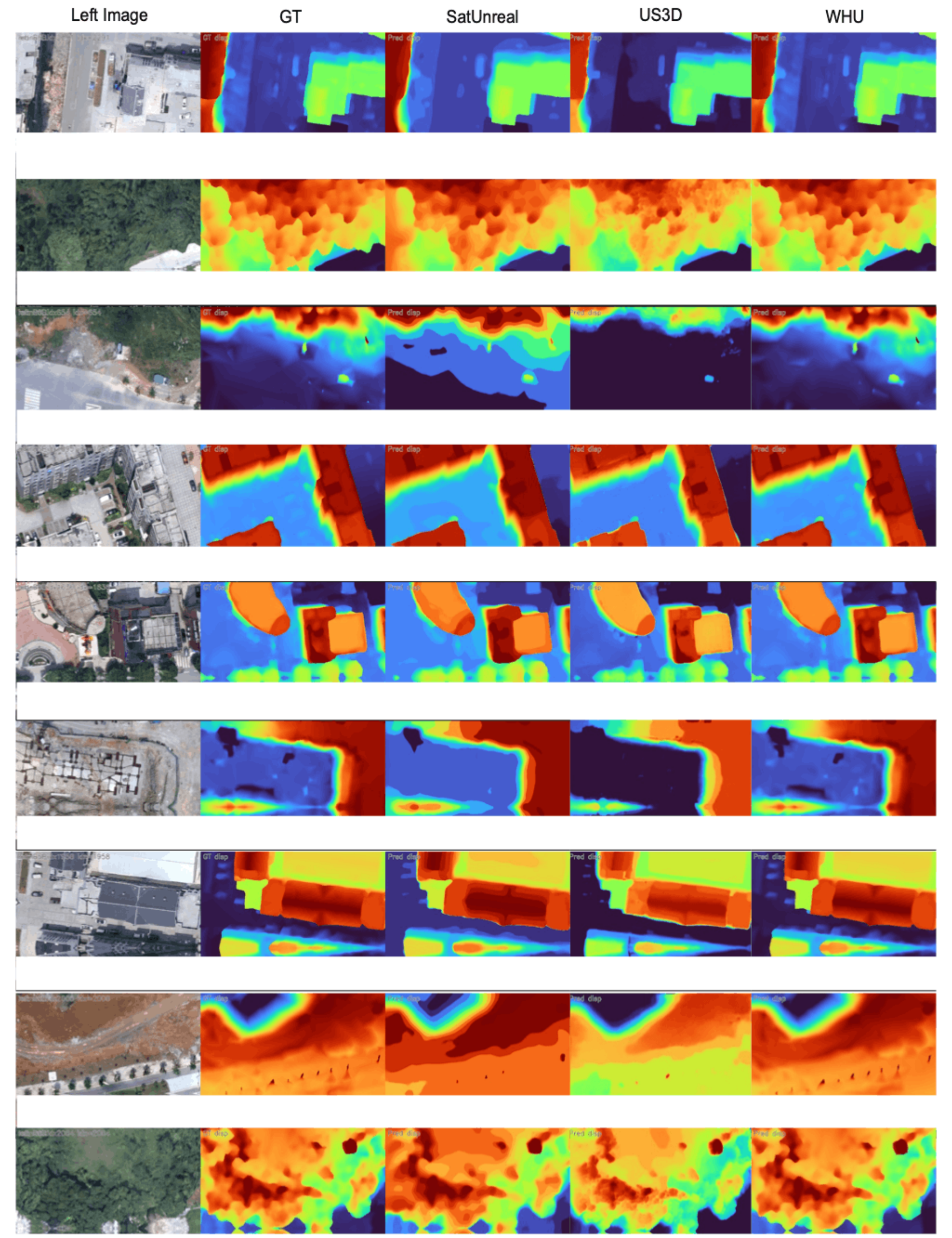

Figure S10. Qualitative evaluation on the WHU-Stereo test set. From left to right: Left image, Ground Truth (GT), and inference results from models trained on SatUnreal, US3D, and WHU-Stereo.

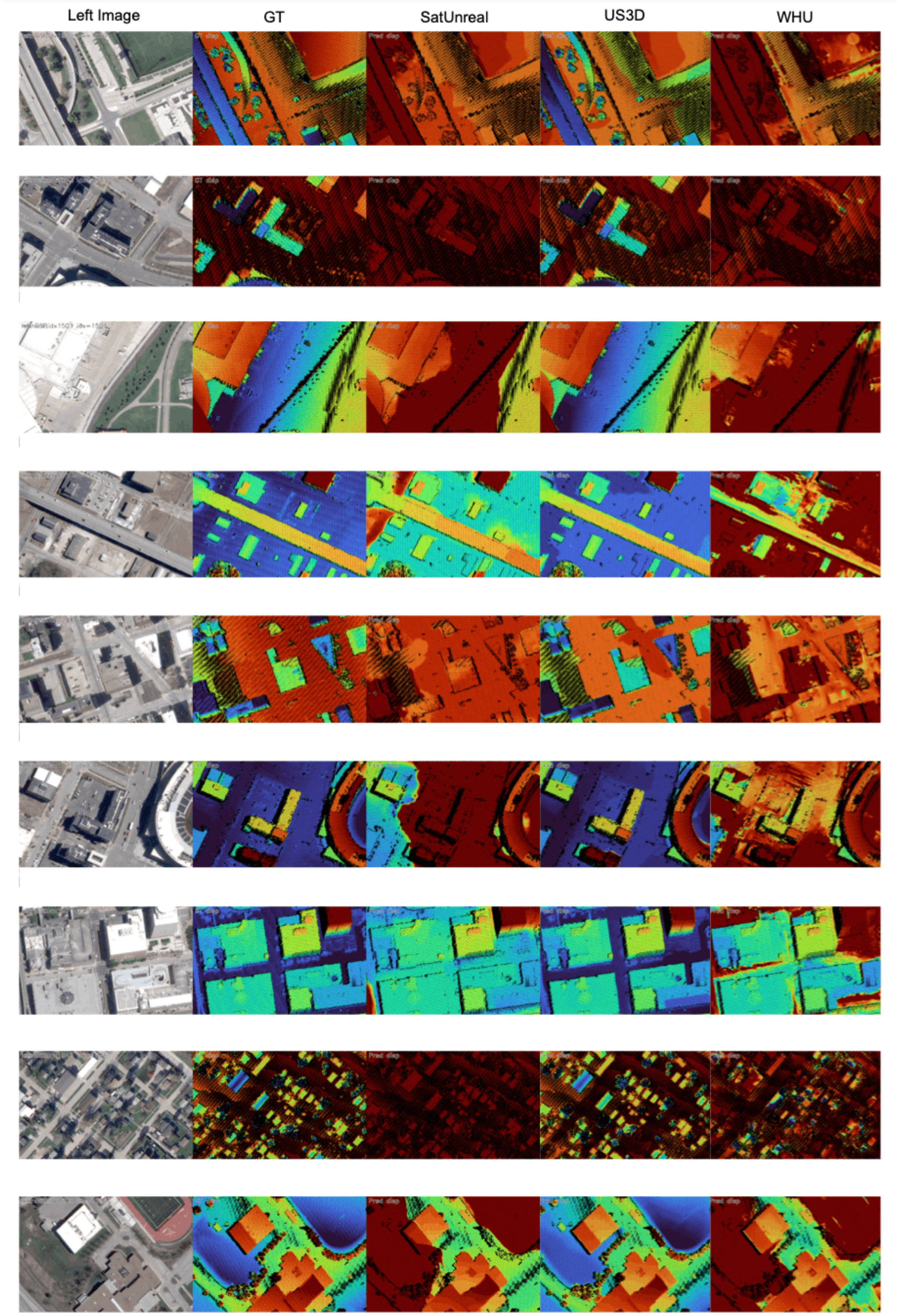

Figure S11. Qualitative evaluation on the US3D test set. From left to right: Left image, Ground Truth (GT), and inference results from models trained on SatUnreal, US3D, and WHU-Stereo.

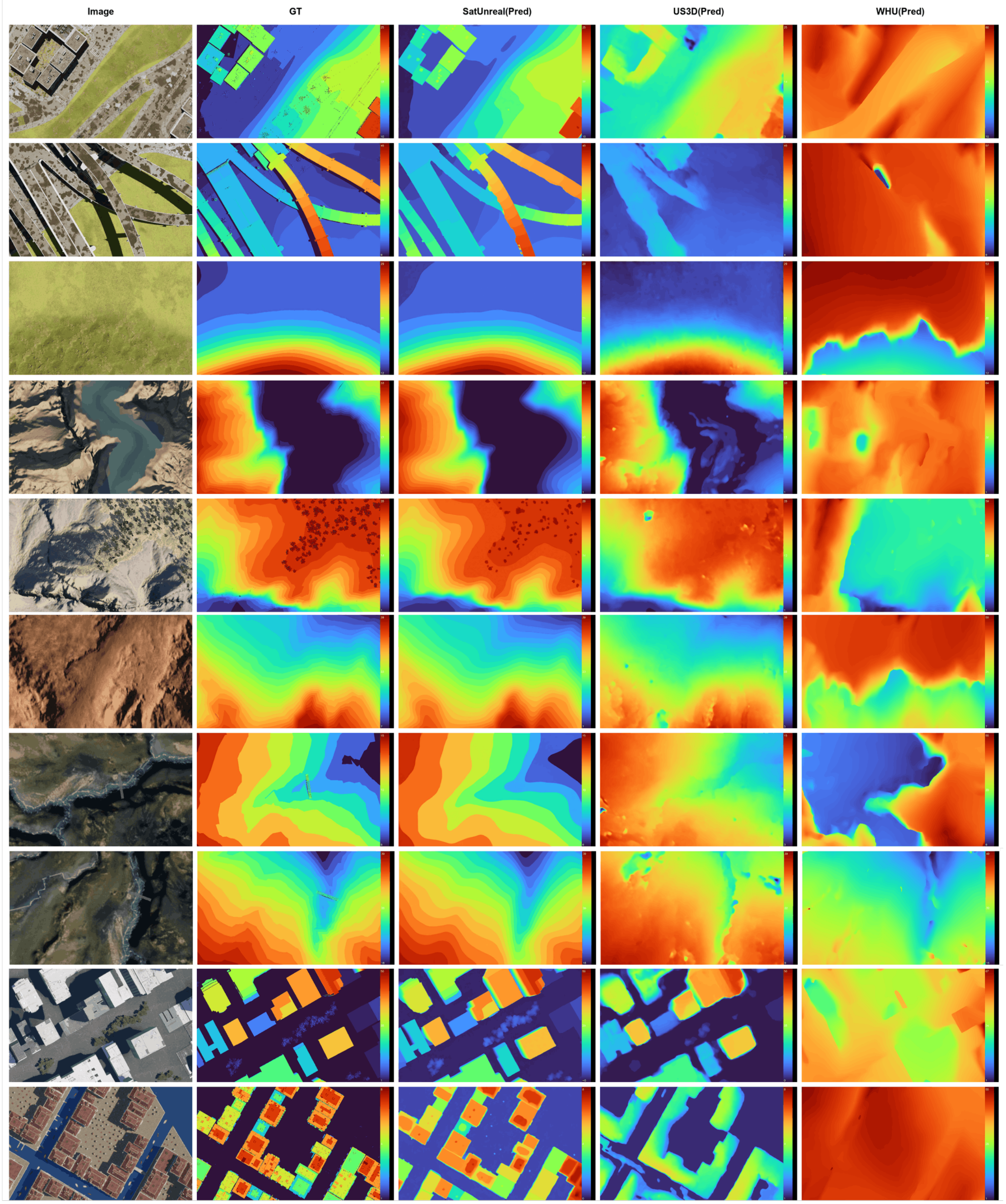

Figure S12. Qualitative evaluation of DLNR on the SatUnreal test set. The SatUnreal-trained model naturally captures the smooth elevation changes and structural layouts, whereas the WHU-trained model struggles to resolve the geometries.

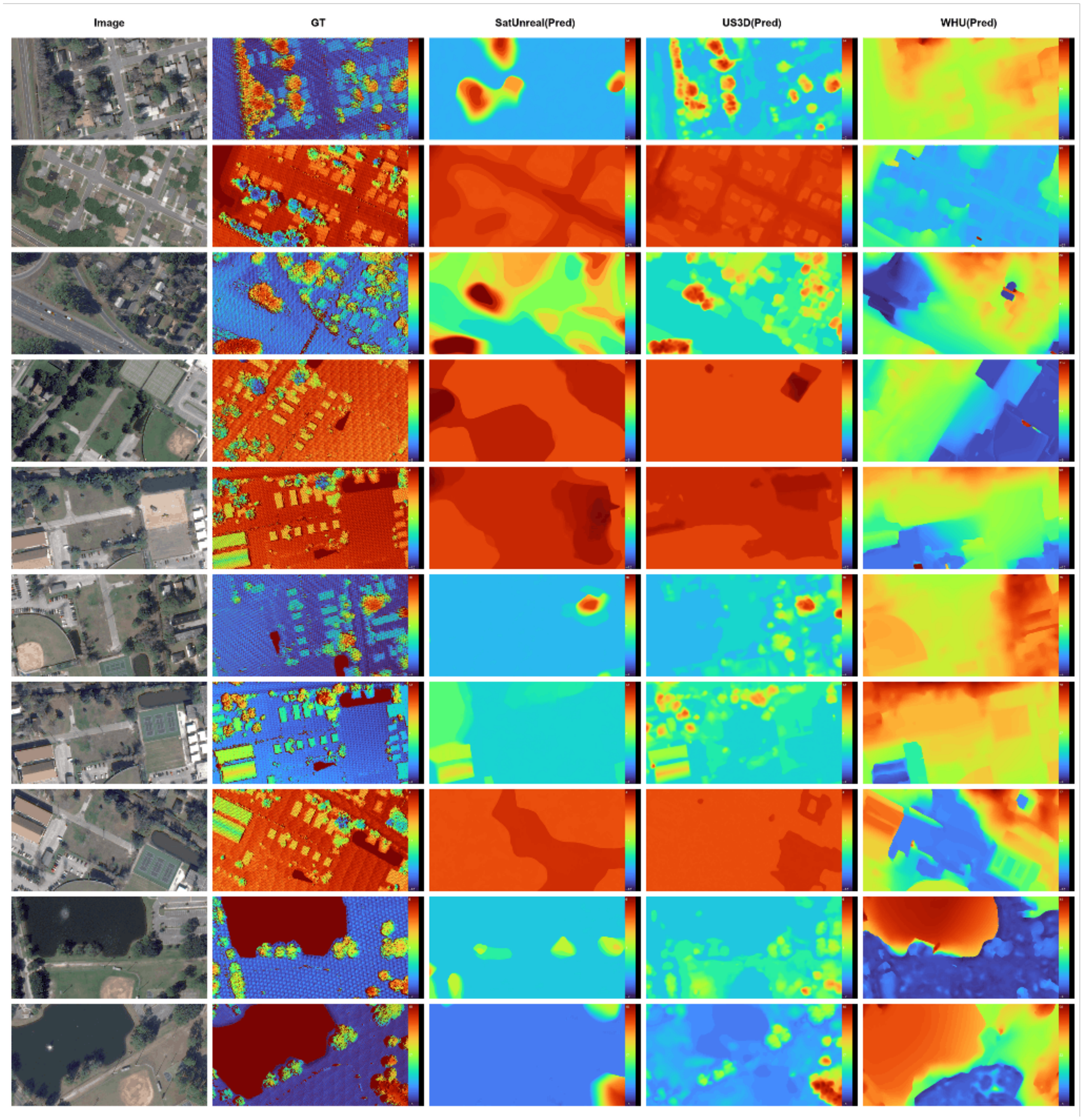

Figure S13. Qualitative evaluation of DLNR on the US3D test set. Due to the fundamental mismatch between DLNR's global smoothing mechanism and US3D's high-frequency LiDAR variations, all models prioritize global smoothness over local micro-structures.

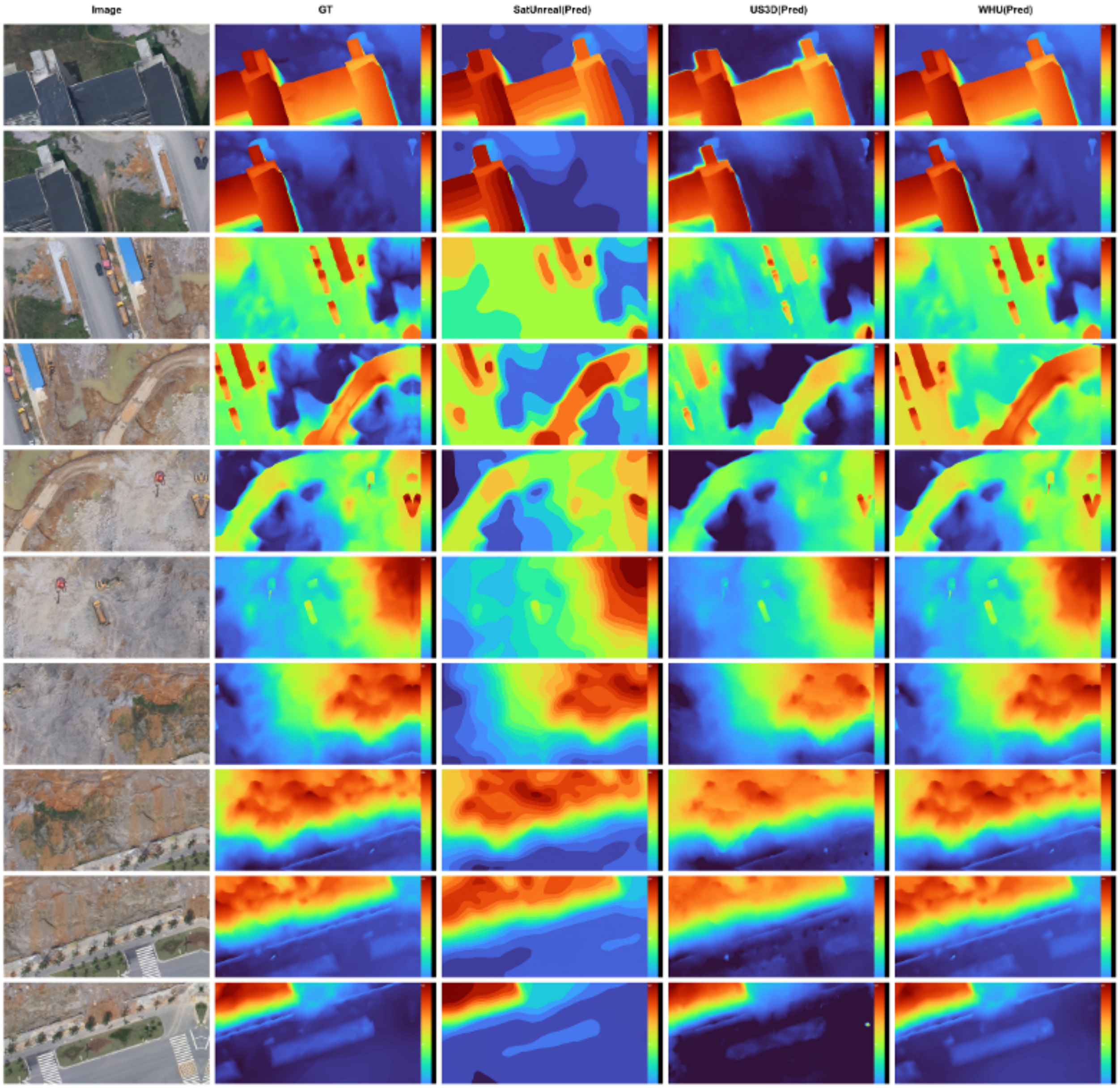

Figure S14. Qualitative evaluation of DLNR on the WHU-Stereo test set. The SatUnreal-trained model successfully generalizes to the real-world WHU domain, delivering results that are highly competitive with, and often smoother than, the in-domain baselines.