GrounDiff: Diffusion-Based Ground Surface Generation from Digital Surface Models

Supplementary Material

This supplementary material provides additional implementation details, extended ablation studies, and qualitative results to complement the main paper. We organize the content as follows:

1. Dataset Details

1.1. ALS2DTM Datasets

The ALS2DTM benchmark datasets [4] consist of **DALES** and **NB**, both with predefined train/validation/test splits. DALES contains 29 training, 10 validation, and 11 test samples, while the NB dataset comprises 84 training, 42 validation, and 42 test samples, each covering a $500\,\text{m} \times 500\,\text{m}$ area at $0.1\,\text{m/px}$ resolution. The DSMs were generated via maximum grid sampling from LiDAR data, while the DTMs were obtained from previous work [4]: the DALES DTMs were acquired from the Canadian governmental geoportal, whereas the NB DTMs were produced through ground classification, manual corrections, and interpolation using TIN. The NB dataset includes a variety of topologies, ranging from urban and suburban areas to forests, whereas DALES is limited to urban scenes. Representative examples of samples from both datasets are shown in Fig. 1.

1.2. USGS (OpenTopology) Datasets

We utilize three regions from the OpenTopology portal, following prior work:

SU (Salt Lake City, Utah). Captured over 2013–2014 using airborne LiDAR, this region covers all of Salt Lake City, totaling approximately 1360 km^2. Due to its large size, SU was divided into three datasets in previous work [1]: SU-I ($\sim$ 7521 m$\times$3871 m), SU-II ($\sim$ 7090 m$\times$3640 m), and SU-III ($\sim$ 6718 m $\times$ 3092 m), all at 0.5 m/px resolution. Each dataset covers approximately 90%, 80%, and 40% urban areas, respectively, with the remainder consisting of mountainous terrain. The three datasets are shown in Figs. 2 to 4.

RT (Refugio, Texas). Acquired by the National Center for Airborne Laser Mapping (NCALM) along the Mission River in Refugio, Texas, following Hurricane Harvey on August 5–6, 2018, using airborne LiDAR, this dataset spans 7196 m × 11883 m at 1 m/px resolution. It covers approximately 90% rural area, including a river and plantation regions (agricultural fields or forested areas) with varying vegetation height, while the remaining 10% is suburban. The region is illustrated in Fig. 5.

KW (Kautz Creek, Washington). Captured on August 28, 2012, within Mount Rainier National Park, Washington, this dataset covers the Kautz Creek watershed (581,000 m × 5,189,000 m) at 1 m/px resolution. It was collected to study landscape response to debris flows and associated hazards. The area features steep mountainous terrain with abrupt elevation changes (alpine) and low-growing vegetation. Fig. 6 provides a visualization of the dataset.

For consistency with our GrounDiff, all datasets are divided into 256 × 256 patches, resulting in 1734 SU-I, 1540 SU-II, 1247 SU-III, 4018 RT, and 1760 KW samples. All DSMs and DTMs were downloaded from the OpenTopography portal.

2. Implementation Details

2.1. Data Preprocessing

For datasets providing only LiDAR point clouds (DALES and NB), we generate DSMs through rasterization by selecting the maximum elevation within each grid cell.

We divide training and validation sets into 256×256 tiles. Our augmentation pipeline includes:

- Random rotations from {0°, 90°, 180°, 270°} with additional jittering within the (-5°, 5°) range.
- Multi-scale resizing to {256×256, 512×512, 1024×1024} to simulate varying metric pixel resolutions.
- Random cropping of a 256×256 tile.
- Horizontal and vertical flipping following [1].

Each augmentation step is applied with 0.5 probability. Augmentation is followed by resizing to 256×256 to match network input requirements.

2.2. Normalization Strategy

We employ min-max normalization computed from valid pixels across both DSM and DTM, mapping all values to the

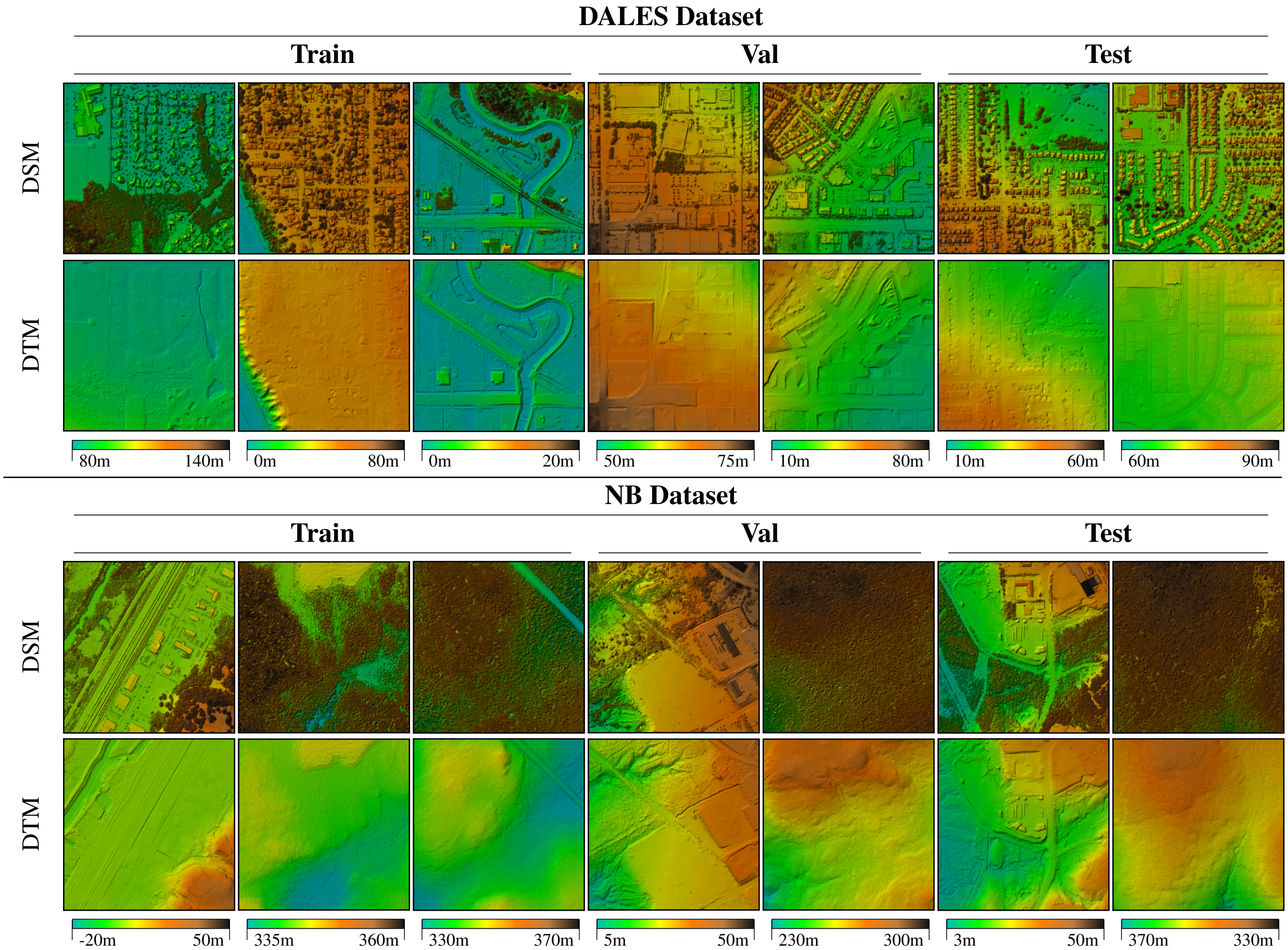


Figure 1. **Representative terrain types from the ALS2DTM benchmark datasets [4]**. Each block shows DSM (top), corresponding DTM (middle), and elevation bars (bottom) for training, validation, and test splits. Top block: DALES dataset; bottom block: NB dataset.

[-1, 1] range while setting invalid regions to zero. Specifically, we calculate the global minimum from both rasters' minima and the global maximum from both rasters' maxima, then apply the transformation:

$$x_{\text{norm}} = 2 \cdot \frac{x - \min(s_m, g_m)}{\max(s_m, g_m) - \min(s_m, g_m)} - 1\,, \qquad (1)$$

where s_m and g_m denote the sets of valid pixels in the DSM s and the DTM g respectively, as defined by the mask m. This approach contrasts with prior methods using global standardization [2] or data localization [1], providing better numerical stability for diffusion processes, as demonstrated in our ablation study. Binary masks m indicating valid pixels undergo identical augmentation transformations to ensure spatial consistency and exclude invalid regions from loss computation.

2.3. Training Configuration

Networks are trained using the AdamW optimizer [5] with learning rate 1e-4, weight decay 0.01, and maximum 1000 epochs with early stopping. A cosine annealing scheduler with 500 warmup steps controls learning rate decay. The diffusion process uses $T = 10$ denoising steps by default unless otherwise specified, with a cosine noise scheduler ranging from 0.0001 to 0.02. Training uses batches of size 16. Loss hyperparameters are empirically set as: $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, $\lambda_\nabla = 0.1$, $\lambda_c = 0.1$.

3. Hardware and Timing Performance

3.1. Hardware Requirements

Our GrounDiff model contains 62.6M parameters and requires approximately 500MB of memory during inference. All training and testing are conducted on NVIDIA A40 GPUs with 48GB VRAM using PyTorch.

3.2. Training Time

Training a single model takes approximately 6 hours to 1 day on an NVIDIA A40 GPU, depending on the dataset

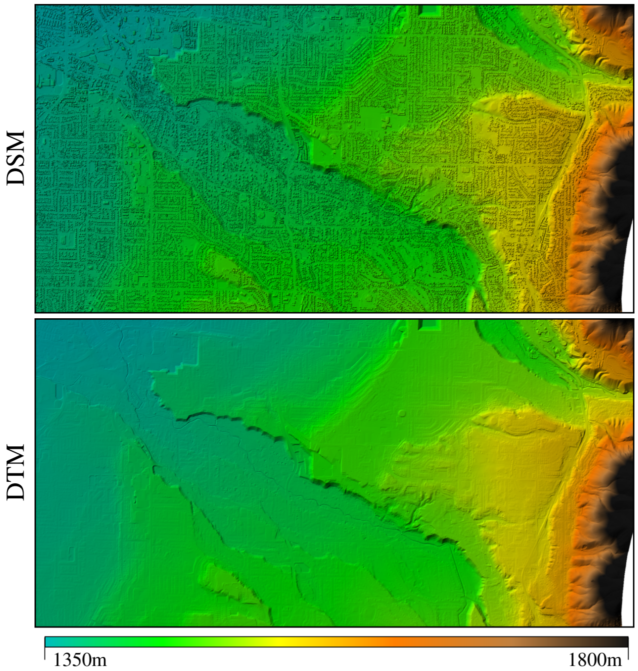


Figure 2. **Visualization of the SU-I dataset [8]**. Top: DSM, middle: DTM, bottom: elevation bar.

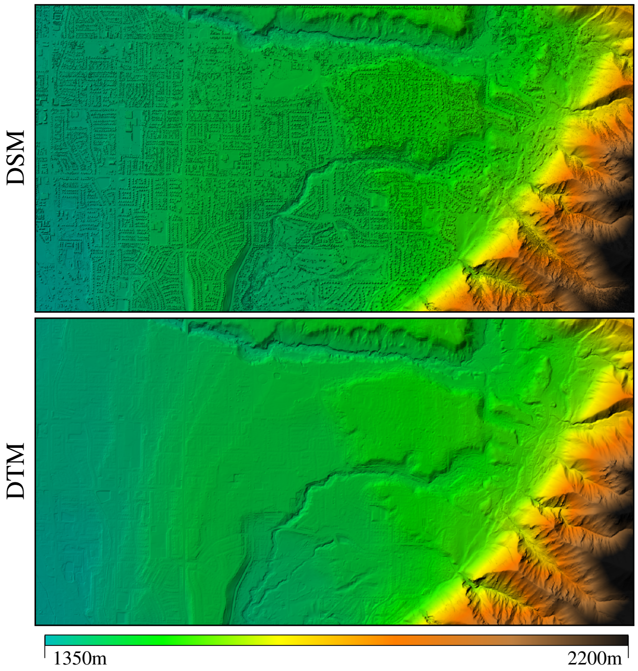


Figure 3. **Visualization of the SU-II dataset [8]**. Top: DSM, middle: DTM, bottom: elevation bar.

and experiment configuration. Larger timestep values require more time as validation steps are slower. Convergence typically occurs within 10K to 20K iterations depending on dataset complexity.

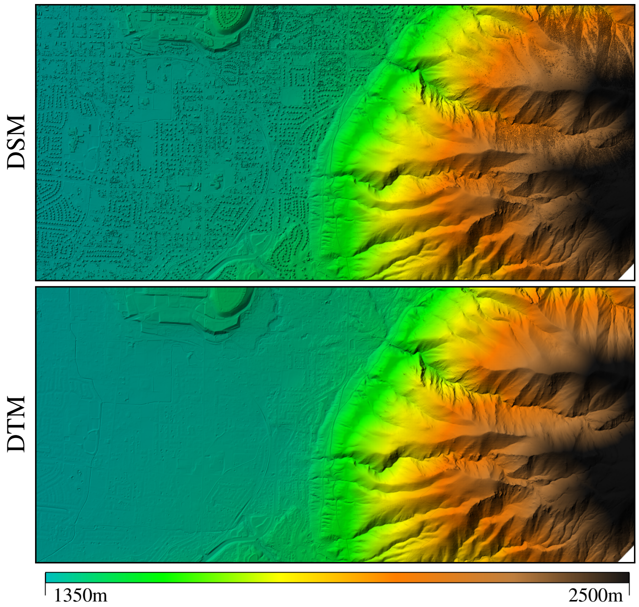


Figure 4. **Visualization of the SU-III dataset [8]**. Top: DSM, middle: DTM, bottom: elevation bar.

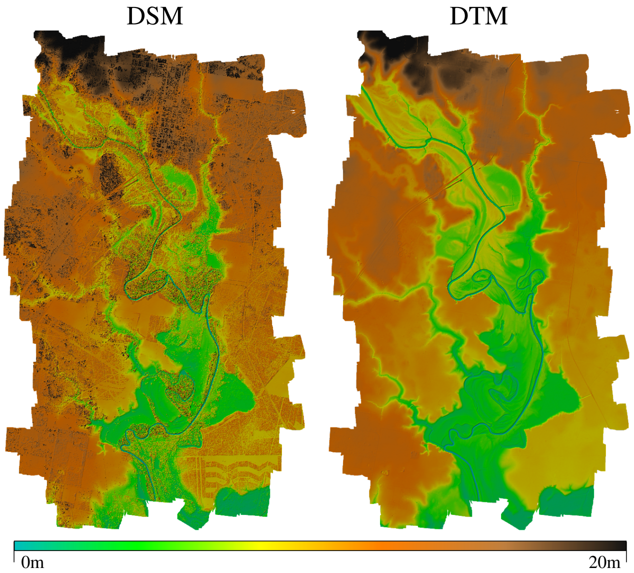


Figure 5. **Visualization of the RT dataset [6] visualization**. Left: DSM, right: DTM, bottom: elevation bar.

3.3. Inference Speed

During inference, a single reverse diffusion step on a 256 × 256 tile takes approximately 60 ms on our GPU. For T diffusion steps, the per-tile inference time is:

$$t_{\text{tile}} = 0.06 \cdot T \quad \text{[s]}. \tag{2}$$

Given an input of width W and height H (in pixels), tile size $P = 256$, and stride S, the number of tiles along each

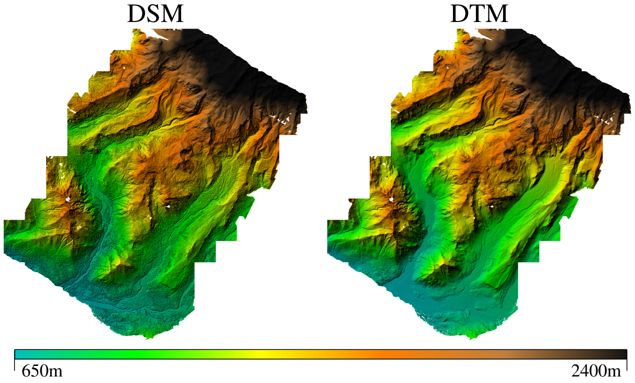


Figure 6. **Visualization of the KW dataset [7].** Left: DSM, right: DTM, bottom: elevation bar.

axis is:

$$N_x = \left\lceil \frac{W-P}{S} \right\rceil + 1, \qquad N_y = \left\lceil \frac{H-P}{S} \right\rceil + 1, \tag{3}$$

yielding a total of:

$$N_{\text{tiles}} = N_x \cdot N_y. \tag{4}$$

The overall inference time becomes:

$$t_{\text{total}} = N_{\text{tiles}} \cdot t_{\text{tile}}. \tag{5}$$

For an area of size A (in km^2) at ground sampling distance r (in m/pixel), the image side length (in pixels) is:

$$W = H = \frac{1000 \cdot \sqrt{A}}{r}, \tag{6}$$

which directly determines N_{tiles} through the equations above. The scaling is approximately linear with area and quadratic with resolution. We show approximate timing examples in Tab. 1.

Area	Resolution	Stride	Time
1 km^2	1.0	256	2
1 km^2	0.5	256	8
1 km^2	0.5	128	32
5 km^2	1.0	256	10
5 km^2	1.0	128	40
10 km^2	1.0	256	20

Table 1. **Processing time for different area sizes, spatial resolutions, and tile strides using our GrounDiff with PrioStitch.** All times are in minutes; resolution is in meters per pixel. All times are reported for $T = 10$.

Compared to a simple divide-and-predict strategy, our PrioStitch strategy introduces minimal overhead: the prior DTM is computed once from a low-resolution version of the input, and blending operations are negligible as they involve straightforward functions. The dominant computational cost arises from per-tile inference, which scales predictably with the number of tiles and thereby with input size and resolution. Note that we do not include batching or multiprocessing strategies in these timing computations, using only single-tile batches during network inference.

4. Analysis of Diffusion Steps

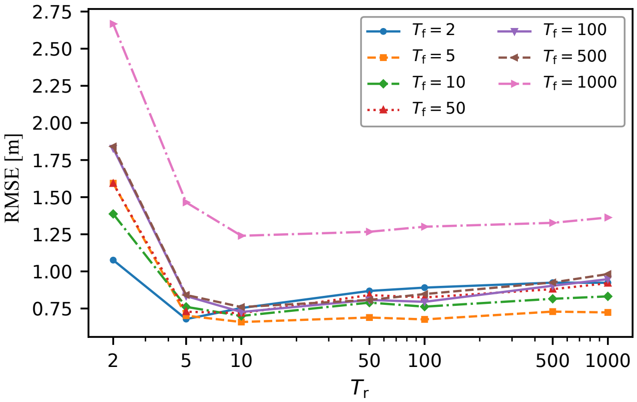


Figure 7. **RMSE performance across different diffusion timesteps.** Models trained with different T_f values are evaluated across varying T_r during testing.

We analyze the impact of diffusion steps on model performance through comprehensive experiments on GeRoD splits defined in the main paper. We evaluate diffusion models trained with different forward timesteps (T_f) across various reverse inference timesteps (T_r). The results in Fig. 7 demonstrate that optimal configurations lie in the moderate timestep range.

Models trained with minimal timesteps ($T_f = 2$) exhibit high instability and poor performance across most inference settings, with RMSE increasing when reverse timesteps exceed the training value. Training with only two timesteps is insufficient for denoising, as the setup approaches a single-pass UNet. Conversely, models trained with extensive timesteps ($T_f = 100, 500, 1000$) suffer from degraded performance and prohibitive computational costs, with $T_f = 1000$ producing extremely high RMSE. We hypothesize that this occurs because the denoiser is expected to handle finer-grained structures, which requires higher network capacity and is inherently more challenging.

Both moderate settings ($T_f = 5, 10$) are stable and show a performance plateau after reaching the training timesteps, indicating that at least this number of steps is needed to reach optimum performance. We adopt $T_f = 10$ as our default configuration because it achieves its lowest RMSE precisely at $T_r = T_f = 10$, providing consistency between training and inference regimes while maintaining computa-

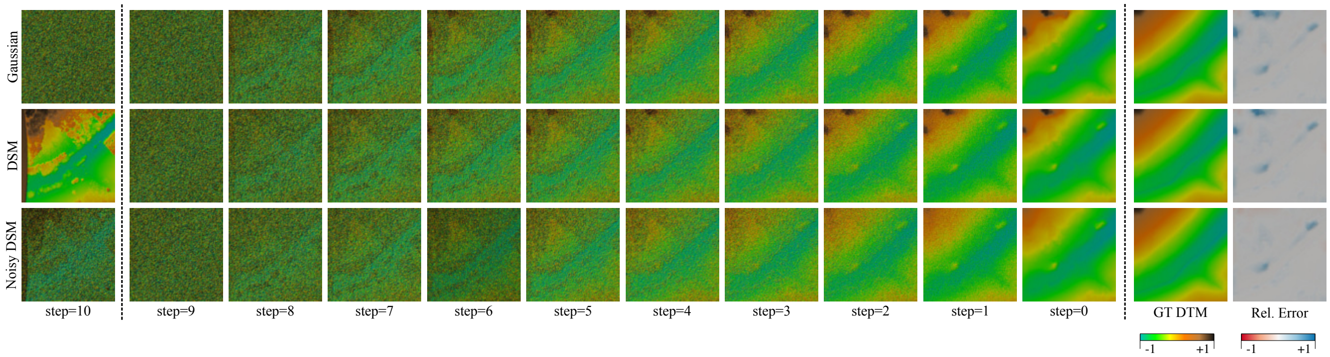


Figure 8. **Diffusion-based denoising progression for $T_f = T_r = 10$.** Progressive generation of cleaner terrain DTMs conditioned on the DSM, starting from Gaussian noise, raw DSM, or noisy DSM. We show the pure denoiser output terrain $s - \hat{r}$ without gating for clear visualization, highlighting the learned interpolation capability. Intermediate steps progress from the initial input (step 10) to the final output (step 0). Errors are color-encoded from red (-1) to blue (+1), and all elevations are normalized to the [-1, 1] range.

tional efficiency.

Fig. 8 visualizes the progressive denoising process, showing how GrounDiff iteratively refines terrain structures for $T_f = T_r = 10$. Initialization with pure Gaussian noise or raw DSM alone results in higher errors than when fusing the DSM with noise. This indicates that adding stochasticity to the DSM introduces structural variations that assist the denoiser in the diffusion process. This also demonstrates how diffusion naturally aligns with the ground filtering task, treating non-terrain elements as noise to be systematically removed.

5. Additional Qualitative Results

5.1. DTM Generation

We provide qualitative results of GrounDiff's performance across diverse environments and challenging scenarios.

Fig. 9 presents a comprehensive overview of our method's performance across all six test datasets. The ground probability maps demonstrate how our model confidently identifies terrain versus above-ground structures, with bright regions indicating high confidence in ground classification. The error maps reveal that most inaccuracies occur beneath buildings and in densely vegetated areas, where true ground measurements are unavailable. In these regions, the ground-truth is typically filled using triangulation-based interpolation. However, our GrounDiff produces physically plausible surface reconstructions that show higher errors, while still better reflecting the actual scene. Importantly, in regions densely covered with vegetation where the ground is nearly invisible (e.g., RT dataset [6]), our method still produces reasonable surface predictions.

These additional results further demonstrate GrounDiff's robustness across diverse environments and its ability to handle challenging scenarios with reasonable performance.

5.2. Road Reconstruction

Fig. 10 provides visual comparison of road surface reconstruction across different scenarios. For urban regions with bridges (first row), FlexRoad [3] can model elevated bridge structures because it uses segmentation-based road extraction and fits NURBS surfaces to identified road segments. In contrast, our GrounDiff is trained to remove all above-ground structures including bridges, making it more accurate at modeling the underlying terrain and tunnel areas beneath bridges while sacrificing elevated road surface representation. Additionally, classification artifacts from ground detection may result in incomplete modeling of road surfaces on bridges. By training our model on data where bridges are part of the DTM, road modeling could be completely handled by our method.

Across all scenes, GrounDiff produces visually more coherent surfaces with structurally plausible terrain continuity. In all cases demonstrates superior road edge modeling compared to FlexRoad [3], with cleaner transitions between road surfaces and adjacent terrain. Our method effectively handles abrupt elevation variations and discontinuities in road surfaces, producing more accurate local topography.

However, the fine-grained mesh details in our reconstructions, while geometrically accurate, result in slightly reduced surface smoothness compared to FlexRoad's mathematically constrained NURBS approach. Our extended version, GrounDiff+, further improves smoothness while remaining flexible, preserving sharp transitions and fine details without significantly compromising precision.

6. Ablations on GrounDiff

We report comprehensive ablation results on the GeRoD dataset [3], including reverse diffusion initialization schemes, loss functions, and normalization strategies. The dataset splits follow the description in the main paper.

The results in Tab. 2 provide a detailed analysis of each design choice. Initializing the reverse diffusion with either

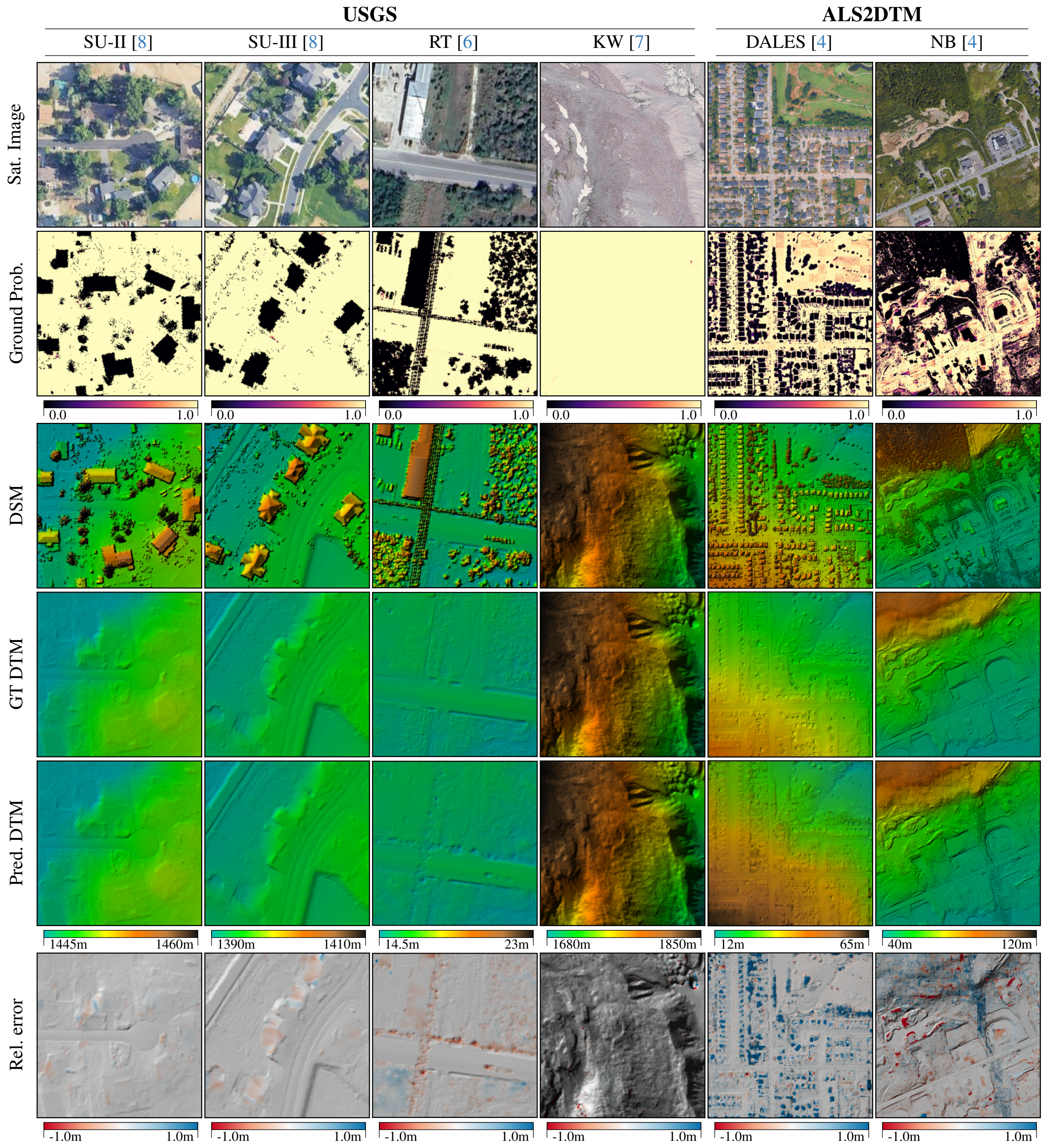


Figure 9. **Ground generation results.** From top to bottom: satellite imagery, ground probability map, input DSM, ground-truth DTM, predicted DTM, and relative error. Examples cover diverse environments: urban regions (SU-II, SU-III [8]), suburban areas (RT [6], NB [4]), steep mountainous terrain with gentle elevation changes (KW [7]), and urban areas (DALES [4]). Satellite imagery is from Google Maps and may not be temporally aligned with the geospatial data due to differences in capture dates.

pure noise or the DSM yields competitive results, with the DSM performing slightly better by providing a structured prior while maintaining stochasticity. Combining stochasticity with the DSM further improves performance, supporting the idea of treating the DSM structure as a form of noise. Using only the $\mathcal{L}_1$ loss increases both RMSE and MAE,

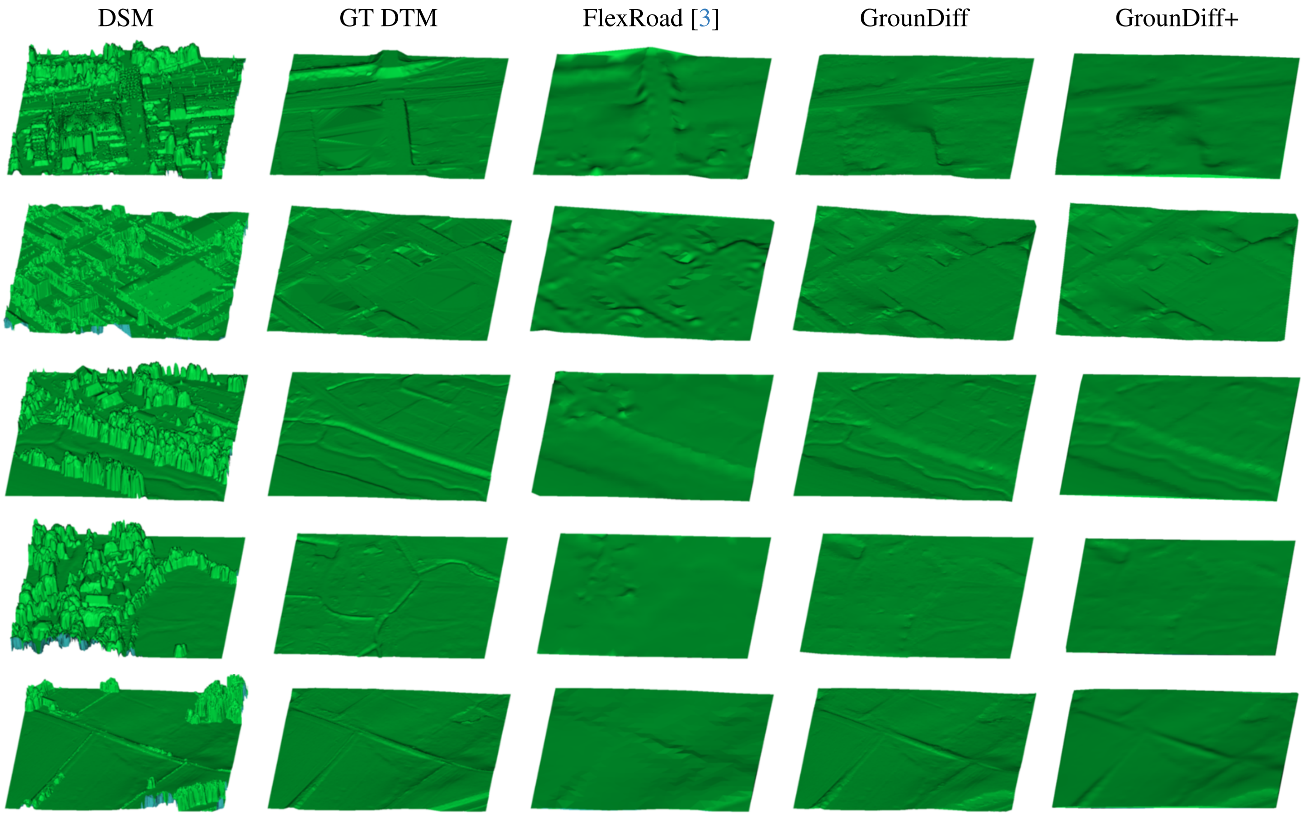


Figure 10. **3D mesh visualizations for road reconstruction from samples in the GeRoD dataset [3].** Each row shows a different scene comparing the input DSM (left), ground-truth DTM, FlexRoad [3], our GrounDiff, and its smoothness-enhanced version GrounDiff+ (right). Our method recovers the underlying terrain more accurately, while GrounDiff+ achieves higher smoothness while maintaining high precision.

Variant	**RMSE↓**	**MAE↓**	E_{T_1}↓	E_{T_2}↓	E_{tot}↓
Init: Noise	0.723	0.401	1.43	1.06	1.11
Init: DSM	0.715	0.400	1.45	1.05	1.13
Loss: $\mathcal{L}_1$	0.742	0.412	1.56	**0.68**	1.01
Loss: $\mathcal{L}_1$ + $\mathcal{L}_2$	0.708	**0.383**	1.31	0.74	**0.93**
Loss: $\mathcal{L}_1$ + $\mathcal{L}_2$ + $\mathcal{L}_\nabla$	0.825	0.439	1.54	0.89	1.11
Norm: Data Localization [1]	7.279	4.692	16.42	17.69	15.80
Norm: Global Standardization [2]	0.950	0.556	**1.03**	2.14	1.37
Baseline (Ours)	**0.700**	0.393	1.43	1.06	1.11

Table 2. **Extended ablation studies of GrounDiff on GeRoD dataset [3].** Includes initialization, loss functions, and normalization.

whereas combining $\mathcal{L}_1 + \mathcal{L}_2$ improves the overall trade-off between height accuracy and classification metrics. Including the gradient loss $\mathcal{L}_\nabla$ without gating slightly increases errors. Regarding normalization, min-max normalization outperforms both global standardization and data localization, demonstrating the benefit of scale-agnostic learning across varied terrain heights.

Collectively, these extended ablations, together with those in the main paper, reinforce the design choices of our baseline method and highlight the components essential for robust terrain reconstruction.

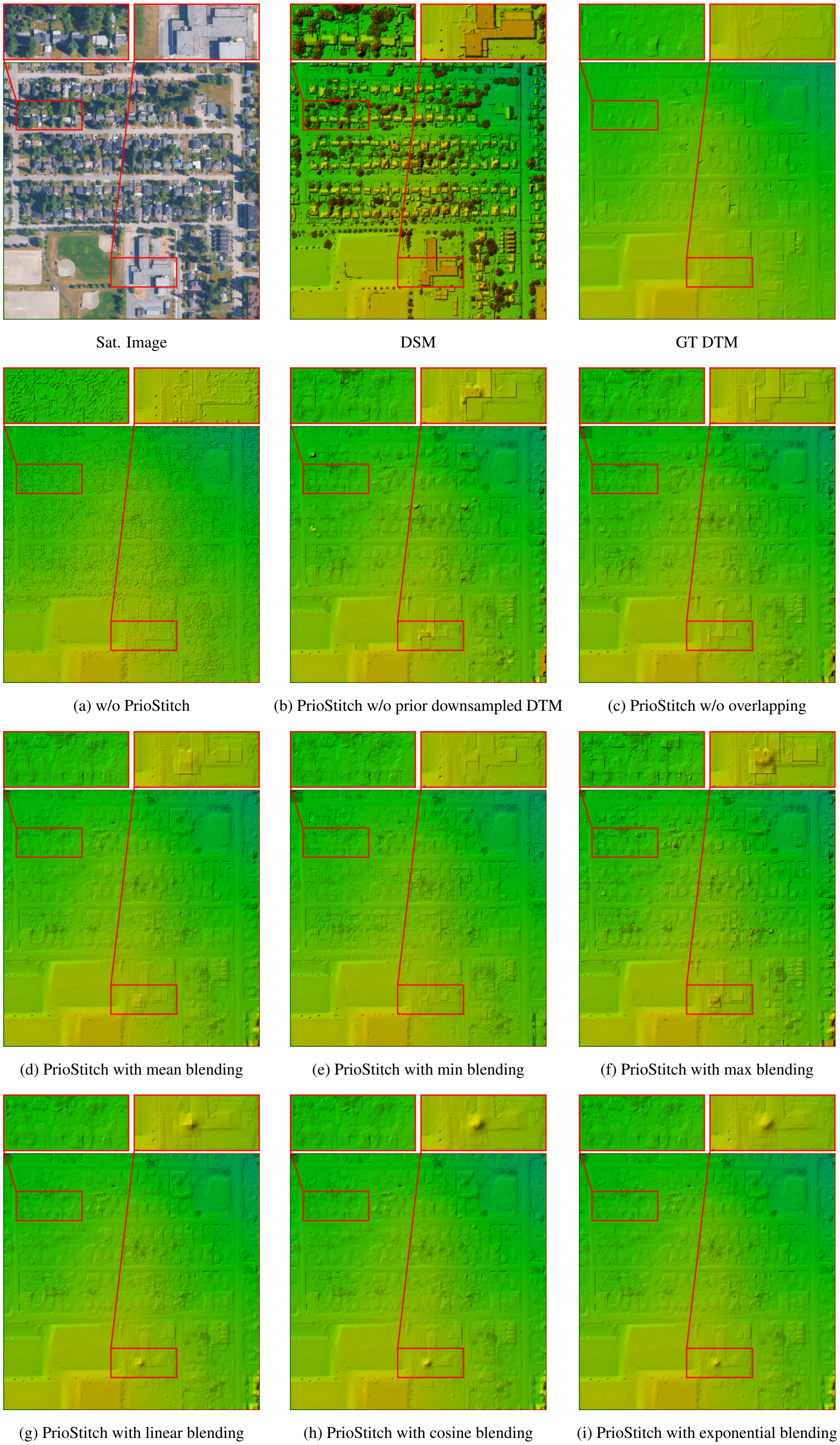
7. Ablations on PrioStitch

We conduct detailed evaluation of our PrioStitch approach on the large-scale urban DALES dataset [4]. Tab. 3 provides quantitative results for different configurations. Fig. 11 provides visual comparison of the different approaches.

7.1. Impact of Global Prior

When PrioStitch is not applied and no prior data is used (a), the network processes a downscaled DSM where downsampling and upsampling introduce interpolation artifacts and eliminate fine-grained details, inducing high regression and classification errors (RMSE=0.780, E_{tot}=17.80%). However, this approach achieves natural smoothness (MAD=3.35°) closest to ground-truth terrain (3.33°) due to the inherent smoothing effect of downsampling that removes small non-ground elements.

Enabling stitching without prior conditioning (b) significantly reduces classification error to 9.31% because the network can observe fine details in full-resolution tiles. However, RMSE increases and surface roughness wors-



Sat. Image

DSM

GT DTM

(a) w/o PrioStitch

(b) PrioStitch w/o prior downsampled DTM

(c) PrioStitch w/o overlapping

(d) PrioStitch with mean blending

(e) PrioStitch with min blending

(f) PrioStitch with max blending

(g) PrioStitch with linear blending

(h) PrioStitch with cosine blending

(i) PrioStitch with exponential blending

Figure 11. **Visual comparison of our PrioStitch ablations.** We show a random large-scale test sample (500 m × 500 m at 0.1 m/pixel) from the DALES [4] dataset. The predicted DTMs (a–i) correspond to our model with the configurations defined in Tab. 3.

	Prio	Stitching	Overlap	Mode	RMSE↓	MAE↓	E_{tot}↓	MAD↓
(a)	✗	✗	✗	-	0.780	0.269	17.80	**3.35**
(b)	✗	✓	✗	-	0.911	0.321	9.31	12.85
(c)	✓	✓	✗	-	0.708	0.256	8.40	13.07
(d)	✓	✓	✓	mean	<u>0.600</u>	0.230	7.68	12.23
(e)	✓	✓	✓	min	**0.514**	**0.196**	**7.63**	12.86
(f)	✓	✓	✓	max	0.941	0.359	9.61	13.37
(g)	✓	✓	✓	linear	0.608	<u>0.224</u>	7.68	<u>11.87</u>
(h)	✓	✓	✓	cosine	0.623	0.226	7.75	12.00
(i)	✓	✓	✓	exp	0.605	<u>0.224</u>	<u>7.67</u>	12.00

Table 3. **Ablation study of the PrioStitch strategy on the DALES dataset [4].** Systematic evaluation of tiling and blending components. **Prior**: whether a low-resolution DTM is used for initialization, otherwise noisy DSM is used; **Stitching**: whether the input DSM is processed in tiles; **Overlap**: whether tiles overlap by 50 percent, stride 128; **Mode**: blending strategy for merging overlapping regions. Metrics include RMSE and MAE in meters, total classification error in percent, and MAD in degrees measuring surface roughness. Ground-truth DTMs have an MAD of 3.33 degrees. **Bold** indicates best performance, <u>underlined</u> indicates second best.

ens (MAD = 12.85°) due to limited contextual information: some tiles contain only vegetation or buildings without visible ground, which challenges accurate terrain prediction.

Incorporating a low-resolution prior DTM auto-generated using GrounDiff (as in configurations (a)) (c) provides essential global context, reducing regression errors (RMSE = 0.708) by 22 % and classification errors (E_{tot} = 8.40 %) by 10 % compared to configuration (b). The prior guides consistent terrain interpretation across ambiguous regions, although surface roughness remains elevated (MAD = 13.07°) compared to the naturally smooth downsampled approach.

7.2. Blending Strategies

We evaluate several blending strategies for merging overlapping tile outputs, as shown in Fig. 12:

- **Mean**: Simple averaging of overlapping regions.
- **Min**: Taking the minimum elevation at each overlap point.
- **Max**: Taking the maximum elevation at each overlap point.
- **Linear**: Linear weighting based on distance from tile edge.
- **Cosine**: Cosine-based weighting for smoother transitions.
- **Exponential**: Exponential decay weighting.

Using overlapping tiles (d-i) further improves all metrics by increasing ground visibility: when individual tiles contain only non-ground regions (vegetation, buildings), overlapping provides additional spatial context where neighboring tiles are more likely to observe ground surfaces.

Minimum blending (e) achieves the best performance across all regression and classification metrics, reducing RMSE and MAE by 14% compared to mean blending (d). This strategy effectively removes residual above-ground artifacts from neighboring tile predictions, as it favors lower elevations that are more likely to represent true ground surfaces. The resulting DTM appears closest to ground-truth visually, though it may introduce sharp elevation jumps at tile boundaries. Conversely, maximum blending (f) performs worst as it preserves above-ground artifacts from overlapping predictions. Continuous blending strategies (linear, cosine, exponential) provide smoother boundary transitions, with linear blending (g) offering the best balance of performance and visual quality.

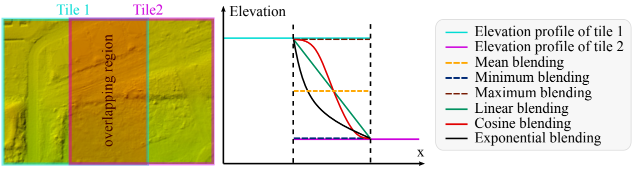


Figure 12. **Blending strategy weighting functions.** Visualization of two overlapping tiles with constant elevation profiles (for simplicity) showing inconsistencies in the overlap region. Different blending modes are applied to merge predictions, demonstrating how each weighting strategy fuses elevation data and handles boundary discontinuities in our PrioStitch approach.

Despite these improvements, the overall MAD remains significantly higher than ground-truth (11.87°-13.07° vs. 3.33°), indicating that some tiles with severely limited ground visibility still produce suboptimal predictions. While prior DTM conditioning provides strong guidance toward correct terrain interpretation, the limited input field of view in challenging scenarios prevents perfect reconstruction of natural surface smoothness.

We encourage further research in this direction by exploring high-resolution DTM generation with networks supporting arbitrary input sizes, as well as point-based diffusion networks to capture more contextual and global information.

8. Limitations

We show examples of failure cases in Fig. 13. All predictions of our GrounDiff are obtained using the model trained on SU-I, where the portion of mountainous and forested regions is very small compared to the overall area, which is predominantly urban. Despite strong performance in urban and suburban regions, our GrounDiff struggles in areas with abrupt elevation changes (e.g., alpine terrain), where sharp elevation gradients resemble those of building facades and are consequently misclassified as non-ground structures, leading to regeneration errors and locally smoothed surfaces. In dense vegetation regions where ground is largely occluded, the model learns to estimate vegetation height but lacks ground reference points in the input data, causing the network to fail completely when elevation differences between pixels are insufficient to identify above-ground struc-

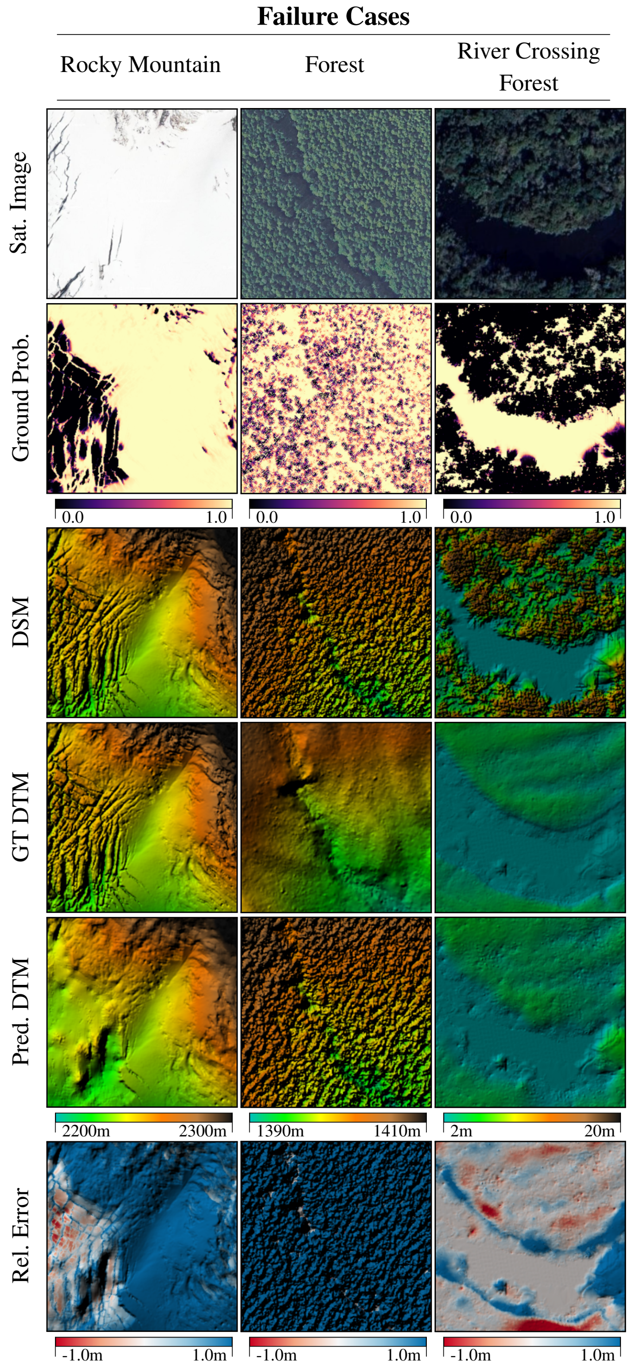


Figure 13. **Failure case examples for ground generation.** From top to bottom: satellite imagery, ground probability map, input DSM, ground-truth DTM, predicted DTM, and relative error. These examples highlight challenging environments: mountainous regions with abrupt elevation jumps, forested areas, and forested regions with rivers. Satellite imagery is from Google Maps and may not be temporally aligned with the geospatial data due to differences in capture dates.

tures. Limited ground visibility can also cause the network to hallucinate terrain. However, when a reasonable number of ground pixels are visible, such as along a river crossing a forest, the surface generation becomes more accurate. Future work could leverage DOP to enrich semantic features or integrate cross-attention within the encoder–decoder architecture. Even in these challenging areas, the generated regions remain visually and physically plausible, and above-ground structures are typically removed successfully.

9. Ethical Considerations

Our approach operates exclusively on elevation data, which contains no personally identifiable information or sensitive geographic metadata. The network processes normalized height values without absolute coordinates, ensuring spatial anonymity. All datasets are publicly available, and the ground sampling distance prevents individual identification.

Training data covers diverse regions; however, performance may degrade in environments substantially different from the training distribution. This limitation is most relevant for extreme topographies underrepresented in current datasets, potentially introducing domain-specific biases.

Our diffusion-based approach is probabilistic and cannot provide deterministic accuracy guarantees. The generative nature of the model may introduce reconstruction artifacts, particularly in occluded regions with limited ground visibility. While extensive validation demonstrates robust performance across benchmarks, we recommend domain-specific evaluation before deployment in safety-critical applications requiring high-precision terrain modeling.

The research scope and data characteristics present no ethical considerations beyond standard machine learning best practices.

References

[1] Hamed Amini Amirkolaee, Hossein Arefi, Mohammad Ahmadlou, and Vinay Raikwar. Dtm extraction from dsm using a multi-scale dtm fusion strategy based on deep learning. *Remote Sensing of Environment*, 274:113014, 2022. 1, 2, 7

[2] Ksenia Bittner, Stefano Zorzi, Thomas Krauß, and Pablo d'Angelo. Dsm2dtm: An end-to-end deep learning approach for digital terrain model generation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10:925–933, 2023. 2, 7

[3] Oussema Dhaouadi, Johannes Meier, Jacques Kaiser, and Daniel Cremers. Shape your ground: Refining road surfaces beyond planar representations. In *2025 IEEE Intelligent Vehicles Symposium (IV)*, pages 2136–2142, 2025. 5, 7

[4] Hoang-An Le, Florent Guiotte, Minh-Tan Pham, Sebastien Lefevre, and Thomas Corpetti. Learning digital terrain models from point clouds: Als2dtm dataset and rasterization-based gan. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:4980–4989, 2022. 1, 2, 6, 7, 8, 9

[5] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 2

[6] OpenTopography National Center for Airborne Laser Mapping (NCALM). Post hurricane harvey mapping of the mission river, texas 2018, 2018. DOI: 10.5069/G9MG7MPD. 3, 5, 6

[7] OpenTopography. Southwest flank of mt. rainier, wa, 2020. DOI: 10.5069/G9PZ56R1. 4, 6

[8] OpenTopography. State of utah acquired lidar data - wasatch front, 2020. DOI: 10.5069/G9TH8JNQ. 3, 6