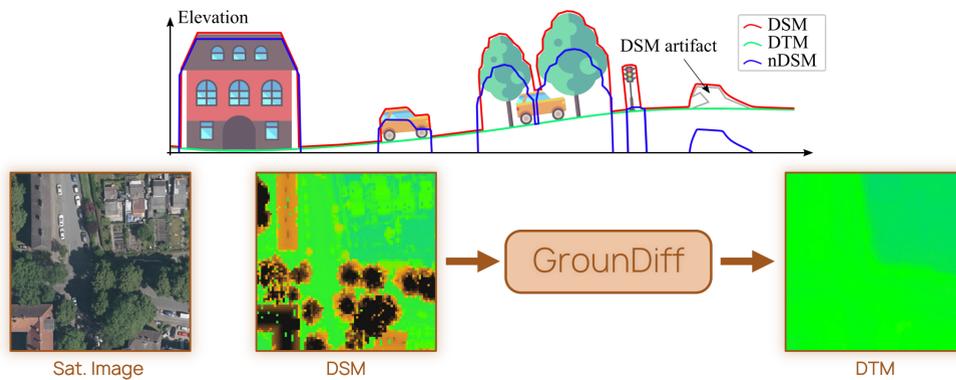


Motivation

Goal

Generate accurate Digital Terrain Models (DTMs), which represent bare-earth elevation, from Digital Surface Models (DSMs) derived from LiDAR or photogrammetry.



Challenge

DSMs contain non-ground objects like vegetation and buildings. Extracting the underlying terrain is difficult in steep, dense, or urban environments.

Why it Matters? DTMs are critical for autonomous navigation, flood modeling, infrastructure planning, and geological analysis.

Previous Work

Traditional Methods

- Morphological:** *PMF*^[1] (Progressive Morphological Filter) and *SMRF*^[2] (Simple Morphological Filter) use structuring elements to remove outlier.
- Statistical & Physical:** *SBM*^[3] (Skewness Balancing) applies statistical thresholds; *CSF*^[4] (Cloth Simulation Filtering) models terrain as a physical cloth.
- Surface-based:** *PTD*^[5] (Progressive TIN Densification) iteratively grows triangular networks.

Learning-Based Methods

- Generative:** *DeepTerRa*^[6] utilizes GANs for DSM-to-DTM translation.
- Regression:** *HDCNN*^[7] employs multi-scale fusion; *RESSUB-Net*^[8] uses residual U-Nets to model elevation differences.

Road Reconstruction Methods (Application Specific)

- Geometric:** planar approximation and *RGT* (Regular Grid Triangulation) for gap interpolation.
- Parametric:** *FlexRoad*^[9] fits NURBS surfaces to segmented road data.

Contributions

- GrounDiff Framework:** The first diffusion-based approach for DTM extraction, treating non-ground structures as noise to be iteratively removed.
- PrioStitch Strategy:** A scalable tiling method that uses a low-resolution global prior to ensure consistency across large geographic areas.
- State-of-the-Art Performance:** Outperforms all baselines, reducing RMSE by up to 93% on ALS2DTM and 47% on USGS.

TL;DR

GrounDiff redefines DSM-to-DTM conversion by "denoising" terrain with diffusion, treating buildings and vegetation as noise and delivering cleaner, more precise ground surfaces than SOTA approaches.

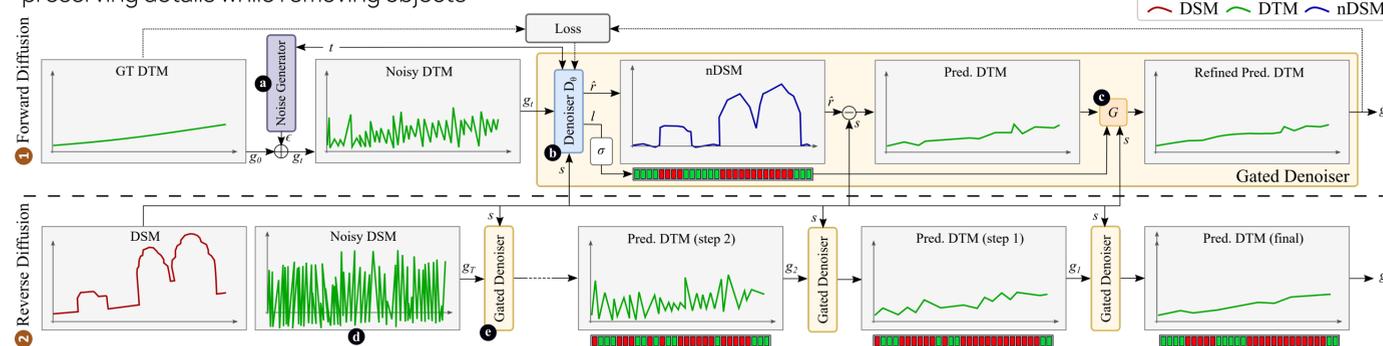
Method

Conditional Diffusion Process

- Forward:** Progressively corrupts the ground truth DTM with noise.
- Reverse:** Iteratively reconstructs the clean DTM from the noisy state, conditioned on the input DSM.

Gated Denoiser Architecture

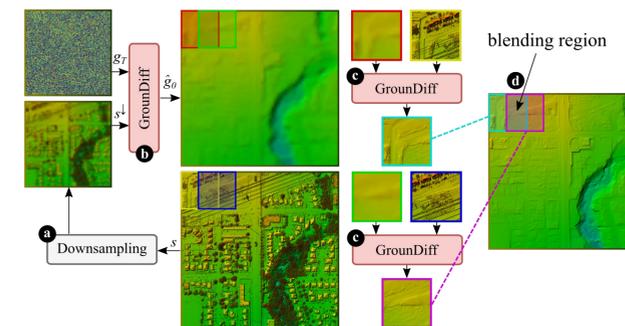
A U-Net backbone predicts a residual correction (nDSM) and ground confidence logits. **Gating Mechanism:** Selectively fuses the input DSM (where ground is confident) with predicted terrain (where occluded), preserving details while removing objects



Architecture (1D for simplicity): **1 Training:** **a** Corrupt DTM g_0 with noise to obtain g_T . **b** DSM conditioned denoiser predicts nDSM and ground logits. **c** DSM - nDSM yields a denoised DTM, which is refined using ground logits to obtain \hat{g}_0 . **2 Inference:** **d** Start from a prior. **e** Iteratively apply DSM conditioned reverse diffusion from $g_T \rightarrow \dots \rightarrow g_0$ to obtain the final DTM.

PrioStitch for Scalability

- Downsample DSM and run GrounDiff to obtain a low resolution prior.
- Tile the original DSM into overlapping patches.
- Condition each patch on the upsampled prior.
- Blend tiles with weighted fusion to produce the final high resolution DTM.



Takeaways

- The diffusion paradigm naturally fits the problem of "denoising" terrain by progressively removing above-ground structures.
- The model generalizes well across diverse terrains without requiring auxiliary data like color or classification masks.
- PrioStitch resolves the trade-off between local detail and global consistency, enabling processing of arbitrarily large areas.

Results

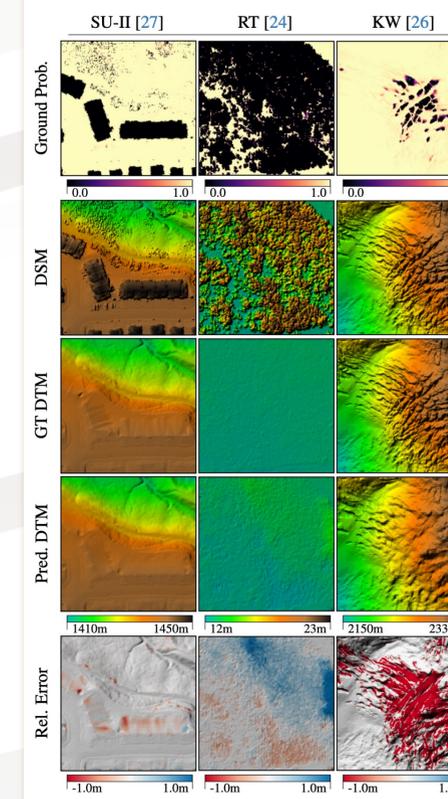
Ground Generation

Datasets: ALS2DTM / USGS.

Quantitative Results:

- vs. Traditional: Achieved 0.51m RMSE on the urban DALES dataset, surpassing CSF (1.19m) and SMRF (4.08m).
- vs. Learning-Based: Reduced RMSE by 92-93% compared to GAN-based *DeepTerRa*^[6] (using only DSM input) and outperformed *HDCNN*^[7] and *RESSUB-Net*^[8] across urban and rural benchmarks.

Qualitative Results:



Road Reconstruction

Dataset: GeRoD.

Quantitative Results:

- Accuracy: Achieved 81% lower Mean Euclidean Distance than the specialized parametric method FlexRoad.
- Smoothness: GrounDiff+ (enhanced variant) improved surface smoothness by 38% while maintaining high accuracy, bridging the gap between pixel-wise accuracy and geometric continuity.

Qualitative Results:

