vehicles

MDPI

Article

Mask-Guided Frequency-Aware Sparse Elevation Map Completion for Off-Road Vehicle Terrain Perception

Lu Zhang [1,2,†], Yuchen Zhang [1,†], Fei Gao [1,*], Yuhang Zhang [1], Chao Li [1] and Xi Li [1]

[1] School of Transportation, Inner Mongolia University, Hohhot 010070, China; imuzhanglu@imu.edu.cn (L.Z.); yuchen_zlg@163.com (Y.Z.); 32524169@mail.imu.edu.cn (Y.Z.); 17636670762@163.com (C.L.); lixi555999@163.com (X.L.)
[2] Inner Mongolia Autonomous Region Intelligent Transportation Vehicle Equipment Engineering Research Center, Hohhot 010070, China
* Correspondence: gaofei-1976@163.com
† These authors contributed equally to this work.

Abstract

Sparse sampling and occlusion limit geometric analysis of LiDAR-derived elevation maps. This study proposes MGFE-Net, a mask-guided and frequency-aware network for elevation-map completion. The method combines elevation, an observation mask, selected-return count, and height variance to represent surface geometry, measurement support, and height dispersion. A pooling-based enhancement module integrates coarse contextual features with learned residual details within a U-Net-like encoder–decoder, while mask-based backfilling retains observed input elevations. Experiments use spatially separated training, validation, and test regions of RELLIS-3D sequence 00000, with exclusion gaps and disjoint source frames. Four network variants are trained from scratch using three random seeds and evaluated against temporally screened multi-frame reference surfaces. MGFE-Net achieves a root mean square error of 0.1593 ± 0.0056 m and a gradient error of 0.1427 ± 0.0024 m/m, reducing the Base U-Net means by 6.77% and 7.95%, respectively. The completed maps increase the computability of slope, step height, and surface roughness, and result in lower mean reference-based feature errors among the evaluated variants. However, exploratory risk evaluation still identifies false-low classifications and spurious connections. These findings support improved reconstruction within the tested spatial split and reference protocol, while broader generalization and safe vehicle traversability remain unverified.

Keywords: off-road terrain perception; LiDAR; elevation-map completion; observation mask; frequency-aware enhancement; geometric terrain features

Check for updates

Academic Editor: Pak Kin Wong

Received: 17 September 2026
Revised: 30 October 2026
Accepted: 3 November 2026
Published: 4 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Off-road vehicles need geometric information to navigate unstructured environments. Slopes, ditches, vegetation, and abrupt height changes cannot be described adequately by obstacle detection alone. Geometry derived from LiDAR, stereo vision, and other range sensors is therefore important for traversability analysis and autonomous navigation [1,2]. Although LiDAR provides direct three-dimensional measurements, sparse sampling and occlusion leave gaps in elevation maps. These gaps also prevent the calculation of terrain features that require neighboring heights, motivating the completion of missing elevations.

A 2.5D elevation map stores a representative height and related attributes in each cell of a horizontal grid. This representation supports terrain-feature calculation, sensor fusion, and navigation. Fankhauser et al. developed robot-centric elevation mapping with

uncertainty estimates [3], and Fankhauser and Hutter introduced the grid map library for multi-layer terrain representation [4]. Yang et al. later proposed neural dense elevation mapping with reconstruction uncertainty estimates [5]. For downstream analysis, however, map coverage and the accuracy of both heights and local height differences must be considered.

Elevation gaps arise from limited vertical sampling, declining return density with range, occlusion, and complex surface geometry. Stölzle et al. studied the reconstruction of occluded terrain elevations using self-supervised learning [6]. Multi-frame accumulation can supply reference observations, but its consistency depends on pose alignment and surface stability. LiDAR-inertial mapping methods such as LIO-SAM address pose estimation [7]; alignment errors and viewpoint-dependent differences can nevertheless remain. Completion experiments must therefore separate input observations from reference-only data and exclude unsupported reference cells from supervision and evaluation.

Conventional approaches include nearest-neighbor interpolation, linear interpolation, and inverse distance weighting. Their performance depends on the available spatial support, surface assumptions, and extrapolation rules. Near a ditch or abrupt height change, a smooth transition may misrepresent the terrain. Completion should therefore be evaluated using elevation errors, local height variation, and reference-based terrain features. Greater continuity or a larger feature-computable area alone does not establish more accurate traversability estimates.

Sparse-depth completion offers related approaches to recovering missing geometric measurements. Uhrig et al. introduced sparsity-invariant convolution to account for measurement locations [8]. Ma and Karaman combined sparse depth with RGB images [9], Eldesokey et al. developed confidence propagation for guided sparse-depth regression [10], and Xu et al. incorporated depth-normal constraints [11]. Surveys by Khan et al. [12] and Hu et al. [13] review the corresponding representations, architectures, objectives, and evaluation methods. These approaches mainly predict depth in the camera image plane. Elevation maps instead represent height over horizontal map coordinates, with a specified height datum and surface-selection rule.

Recent methods also investigate how sparse observations should be processed before reconstruction. BP-Net propagates sparse depth early and combines it with multimodal fusion and refinement [14]. SigNet first predicts coarse dense depth, then refines it using degradation-aware image guidance [15]. Both use color imagery alongside sparse depth. Their treatment of sparse support and spatial context is relevant to elevation completion, although their inputs and prediction coordinates differ from those of a LiDAR-only elevation grid.

U-Net provides a backbone that combines contextual encoding with spatially resolved decoding and skip connections [16]. In image inpainting, partial convolutions explicitly use observation masks [17], LaMa uses Fourier convolutions to provide broad context for large gaps [18], and focal frequency loss weights reconstruction discrepancies in the frequency domain [19]. These studies motivate the use of observation information and features at multiple spatial scales. Our enhancement module combines average pooling with a learned residual-detail branch.

We investigate two questions. First, do an explicit observation mask and multiscale enhancement improve reconstruction when return-count and height-dispersion information are supplied under a common training protocol? Second, do lower reconstruction errors correspond to more accurate terrain features and fewer adverse connections under geometric risk criteria?

MGFE-Net combines elevation, an observation mask, selected-return count, and height variance in a four-channel input. A pooling-based module enhances the bottleneck of a

U-Net-like encoder–decoder, and exact backfilling retains observed elevations. We evaluate four architectural variants using three training seeds and spatially separated regions of RELLIS-3D sequence 00000 [20]. The evaluation covers reconstruction accuracy, performance under different missingness conditions, terrain-feature errors, and geometric risk connectivity.

2. Methodology

MGFE-Net estimates missing elevations while retaining observed input heights. The method consists of four-channel map construction, encoder–decoder prediction with bottleneck enhancement, and observation backfilling. All input channels use selected returns from one anchor scan. Additional scans supply reference targets only for supervision and evaluation. The original observation mask continues to distinguish measured from inferred elevations in the completed map.

2.1. Overview of the Proposed Method

Figure 1 shows the workflow. A single anchor scan is filtered to construct four aligned channels: elevation, observation mask, selected-return count, and height variance. These represent surface height, observation availability, measurement support, and within-cell height dispersion. MGFE-Net predicts a dense elevation map from the normalized channels. Predictions are converted to meters before the original observed elevations are copied into the completed map.

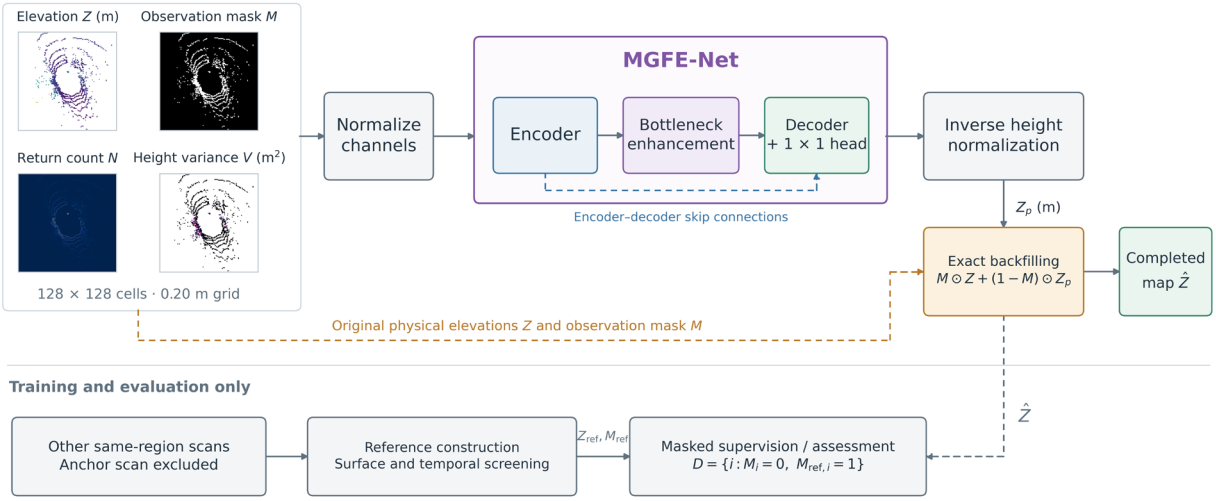

Figure 1. Overall workflow of the proposed MGFE-Net. Gray arrows indicate data flow; the dashed gray connection links the completed map to supervision and assessment during training and evaluation. Blue dashed arrows indicate encoder–decoder skip connections. Orange dashed arrows carry the original physical elevations and observation mask to the backfilling operation.

The four-channel representation is defined as:

$$X = [Z, M, N, V] \in \mathbb{R}^{4\times H\times W} \tag{1}$$

Here, X denotes the four-channel map representation; H and W are the numbers of grid rows and columns; and Z, M, N, and V denote the elevation, observation-mask, selected-return-count, and height-variance channels, respectively. In this study, $H = W = 128$. The observation mask uses $M = 1$ for an observed input cell and $M = 0$ for a missing input cell.

The physical-height prediction is expressed as follows:

$$Z_p = T_Z^{-1}[f_\theta(T_X(X))] \tag{2}$$

Here, f_θ denotes the network with trainable parameters θ, T_X is the channel-normalization operation, and the inverse height transformation converts the predicted normalized heights back into meters. The normalization parameters are estimated exclusively from training inputs and remain fixed for validation and testing.

The network predicts a height for every cell, but predictions replace input values only at missing cells. Observed cells retain their original selected-surface heights. A separate mask defines reference validity. Neither input missingness nor reference validity is inferred from stored numerical values, because zero can also represent a valid height.

2.2. Four-Channel Elevation-Map Representation

Each input uses one Ouster OS1 64-channel scan, subsampled to native ring IDs 3, 7, 11, ..., 63 before filtering to obtain a controlled 16-ring input. Returns are restricted to a horizontal sensor range of 0.5–25 m and sensor-frame heights of −3 to 3 m. The supplied pose transforms them to world coordinates. They are then clipped to their assigned spatial region before surface filtering.

Surface selection uses cloth simulation filtering (CSF) [21] followed by a lowest-supported-layer rule. CSF uses a cloth resolution of 0.20 m, a classification threshold of 0.20 m, rigidness 1, an iteration limit of 500, a time step of 0.65, and slope postprocessing. Retained returns are grouped into 0.20 m world-XY bins. Sorted heights are divided into layers wherever consecutive values differ by more than 0.15 m. If several layers occur, the lowest is retained only if it has at least three returns; otherwise, the bin is rejected. A single layer proceeds to the cell-validity checks. This branch uses neither semantic labels nor reference-only observations.

Let P_i denote the selected input returns assigned to grid cell i, n_i their number, and z_k the relative height of return k. The minimum support threshold is $n_{\min} = 3$. For $n_i > 0$, the mean height and population height variance are calculated as:

$$\mu_i = \frac{1}{n_i} \sum_{k \in P_i} z_k, \quad {s_i}^2 = \frac{1}{n_i} \sum_{k \in P_i} (z_k - \mu_i)^2 \tag{3}$$

The observation mask requires both sufficient support and consistent selected heights:

$$M_i = \begin{cases} 1, & n_i \geq n_{\min} \text{ and } s_i \leq 0.20 \text{ m} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The elevation channel is then defined as:

$$Z_i = \begin{cases} \mu_i, & M_i = 1 \\ 0, & M_i = 0 \end{cases} \tag{5}$$

The selected-return-count and height-variance channels are:

$$N_i = n_i, \quad V_i = \begin{cases} {s_i}^2, & M_i = 1 \\ 0, & M_i = 0 \end{cases} \tag{6}$$

The population variance uses n_i as its divisor. Its units are m^2, whereas N is a return count per grid cell. With a grid resolution of 0.20 m, each cell covers 0.04 m^2. N is retained even when $M = 0$, because a cell may contain selected returns but fail the minimum-support or height-consistency requirement. Thus, a missing cell does not necessarily have a zero return count.

The grid is aligned with the supplied world coordinate system: columns increase along world x and rows along world y. Each 128 × 128 patch covers 25.6 × 25.6 m, with its origin located at the anchor-pose XY translation minus (12.8, 12.8) m. Stored heights are measured relative to the anchor-pose Z translation rather than an absolute geodetic

datum. The supplied poses are used without independent refinement or propagation of pose uncertainty.

Normalization uses the 129 training inputs only. Observed elevations are standardized with a mean of −1.0132655494 m and a standard deviation of 0.2325955095 m. Selected-return counts and height variances are divided by their respective training-input maxima of 86 and 0.0399489701 m^2. Missing elevations remain zero, and the observation mask is unchanged. No normalization clipping is applied. Reference-valid heights use the same height transformation. The accompanying Reproducibility Archive stores the full-precision parameters.

The mask distinguishes missing measurements from valid physical or normalized zeros. Count and variance describe input support and dispersion; neither is calibrated predictive uncertainty. The selected-surface mean also does not establish traversable ground. Vegetation, mixed surfaces, and pose-alignment errors can remain after filtering.

2.3. Network Architecture

The architecture of MGFE-Net is shown in Figure 2. The network follows a U-Net-like encoder–decoder structure with a base feature width of 64. An initial DoubleConv block maps the four-channel input to 64 feature channels. Each DoubleConv contains two 3 × 3 convolutions, each followed by batch normalization and a rectified linear unit (ReLU). Four downsampling stages apply 2 × 2 max-pooling followed by DoubleConv blocks, producing feature widths of 128, 256, 512, and 512.

For a 128 × 128 input, the encoder produces the bottleneck feature:

$$\mathbf{F} \in \mathbb{R}^{C_b \times H_b \times W_b}, \quad C_b = 512, \quad H_b = W_b = 8 \tag{7}$$

The frequency-aware enhancement module operates at this bottleneck. Its pooled branch reduces the spatial dimensions from 8 × 8 to 4 × 4 through adaptive average pooling. Two 1 × 1 convolutions with an intervening ReLU transform the pooled features, which are then bilinearly upsampled to the original bottleneck resolution:

$$\mathbf{F}_l = U[C_{l,2}(\delta[C_{l,1}(P(\mathbf{F}))])] \tag{8}$$

Here, P denotes adaptive average pooling to 4 × 4, $C_{l,1}$ and $C_{l,2}$ denote the two 1 × 1 convolutions, δ denotes ReLU, and U denotes bilinear upsampling to 8 × 8. The pooled branch supplies contextual features at a coarser spatial scale.

The residual-detail input is obtained by subtracting the upsampled pooled representation from the original bottleneck feature:

$$\mathbf{F}_r = \mathbf{F} - \mathbf{F}_l \tag{9}$$

A 3 × 3 convolution followed by batch normalization and ReLU transforms this residual:

$$\mathbf{F}_h = \delta[\mathrm{BN}(C_{3\times3}(\mathbf{F}_r))] \tag{10}$$

The pooled and residual-detail features are concatenated along the channel dimension and fused using a 1 × 1 convolution, batch normalization, and ReLU:

$$\mathbf{F}_f = \delta[\mathrm{BN}(C_f[\mathrm{Concat}(\mathbf{F}_l,\ \mathbf{F}_h)])] \tag{11}$$

Here, C_f denotes the 1 × 1 fusion convolution. A residual connection produces the enhanced bottleneck feature:

$$\mathbf{F}' = \mathbf{F} + \mathbf{F}_f \tag{12}$$

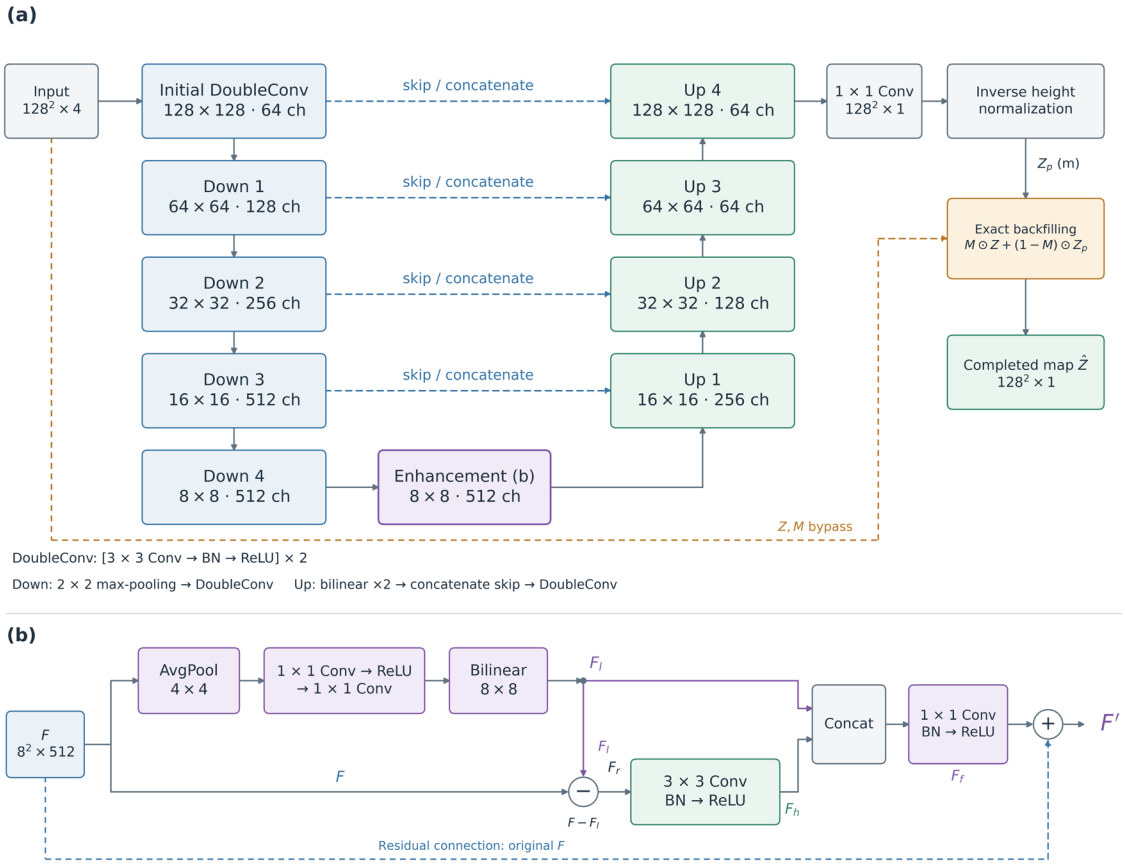

Figure 2. Architecture of MGFE-Net: (**a**) the encoder–decoder with exact observation backfilling; (**b**) the pooling-based feature-enhancement module at the bottleneck. Gray arrows indicate feature flow, blue dashed arrows indicate skip or residual connections, and orange dashed arrows indicate the bypass of the original elevations and observation mask for backfilling. Purple arrows in (**b**) carry the pooled-context feature.

The pooled branch represents spatial context at a coarser scale, while the residual branch processes deviations from the upsampled pooled representation. The terms low-frequency and high-frequency refer to these spatial scales; the module operates through pooling, interpolation, and learned convolutions.

The decoder has four stages. Each bilinearly upsamples its input, concatenates the corresponding encoder features, and applies a DoubleConv. Output widths are 256, 128, 64, and 64 channels, at resolutions of 16 × 16, 32 × 32, 64 × 64, and 128 × 128. Bilinear upsampling uses align_corners = True. A final 1 × 1 convolution predicts one normalized elevation channel.

2.4. Backfilling Constraint and Loss Function

Mask-based backfilling prevents network estimates from replacing observed input elevations. Let Z denote the physical input elevation map, Z_p the network prediction converted into meters, and $\hat{Z}$ the completed elevation map. Backfilling is defined as:

$$\hat{Z} = M \odot Z + (1 - M) \odot Z_p \tag{13}$$

Here, $\odot$ denotes element-wise multiplication. At $M = 1$, the input elevation is retained; at $M = 0$, the predicted elevation is used. This operation is applied consistently to all compared network variants. Omitting the explicit mask channel from a network input does not remove the mask from backfilling or loss-domain construction.

A separate mask, M_{ref}, identifies reference-valid cells. The reconstruction domain D contains only cells for which $M = 0$ and $M_{\text{ref}} = 1$. Reference-invalid cells do not contribute to reconstruction supervision. Within a mini-batch, eligible cells retain their sample identities, and no spatial relationships are constructed between different samples.

The active training objective combines reconstruction and gradient losses:

$$\mathcal{L} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{grad}}\mathcal{L}_{\text{grad}}, \quad \lambda_{\text{rec}} = 1.0, \quad \lambda_{\text{grad}} = 0.5 \tag{14}$$

The reconstruction loss is the mean squared physical-height error over D:

$$\mathcal{L}_{\text{rec}} = \frac{1}{|D|}\sum_{i \in D} \left(\hat{Z}_i - Z_{\text{ref},i}\right)^2 \tag{15}$$

Here, Z_{ref} denotes the reference elevation map and $|D|$ is the number of eligible cells in the mini-batch. This term has units of m^2. A mini-batch with an empty reconstruction domain is rejected rather than evaluated using invalid targets or observed-cell substitutions.

The gradient term uses horizontal and vertical adjacent-cell pairs. An edge is eligible only when both endpoints are reference-valid and at least one endpoint belongs to D. Let E denote the set of eligible edges, with each edge counted once. The gradient loss is:

$$\mathcal{L}_{\text{grad}} = \frac{1}{|E|}\sum_{(i,j) \in E} \frac{\left|(\hat{Z}_j - \hat{Z}_i) - (Z_{\text{ref},j} - Z_{\text{ref},i})\right|}{\Delta}, \quad \Delta = 0.20 \text{ m} \tag{16}$$

This loss measures discrepancies in height increments per unit horizontal distance, in m/m. Eligible edges include missing-to-missing interiors and missing-to-observed boundaries. Edges crossing reference-invalid cells, crop borders, or sample boundaries are excluded. An empty training edge set contributes zero gradient loss; an evaluation metric with an empty denominator is undefined.

The loss terms are evaluated using physical heights, either after inverse normalization or by equivalent rescaling of normalized height differences. The gradient term encourages agreement between predicted and reference height increments. No smoothness regularizer is applied (Table 1).

Table 1. Data-generation and training settings. Spatial bounds are shown to six decimal places. The split manifest in the Reproducibility Archive retains full precision and the source-frame lists.

Item	Setting
Dataset and input	RELLIS-3D sequence 00000; one Ouster OS1 anchor scan; native rings 3, 7, …, 63
Grid and footprint	0.20 m; 128 × 128 cells; 25.6 × 25.6 m
Training world-y band	$[-202.083000, -151.894682)$ m; 129 patches
Validation world-y band	$[-126.294682, -76.106365)$ m; 23 patches
Test world-y band	$[-50.506365, -0.318047)$ m; 27 patches
Spatial exclusion and sampling	25.6 m between bands; at least 1 m between retained anchor centers within a split
Input validity	$n_{\min} = 3$; selected-surface height SD ≤ 0.20 m
Reference support	At most 160 other frames; offset ± 160, step 2; at least 6 returns from at least 3 frames per valid cell
Reference consistency	Pooled point-height SD and frame-mean IQR ≤ 0.15 m; both temporal halves valid; median difference ≤ 0.15 m
Models and random seeds	Base U-Net, Mask U-Net, Freq U-Net, MGFE-Net; seeds 42, 101, 202
Optimization	Adam; initial learning rate 10^{-4}; batch size 8; maximum 200 epochs; cosine annealing

Table 1. *Cont.*

Item	Setting
Checkpoint selection	Minimum pooled validation RMSE in meters; early-stopping patience 20 epochs
Loss weights	Reconstruction MSE: 1.0; gradient L1: 0.5; smoothness: 0

Exact backfilling preserves observed center heights, but completed neighbors can still change slope, step height, and roughness. We therefore evaluate height reconstruction, boundary-gradient agreement, and downstream geometry separately.

3. Experimental Setup

3.1. Dataset and Sample Construction

We use Ouster OS1 64-channel scans and supplied poses from RELLIS-3D sequence 00000 [20]. Spatial regions are fixed before inputs or references are generated. Training, validation, and test regions occupy non-overlapping spatial bands in the supplied world coordinate system (Table 1), separated by 25.6 m gaps. Each gap equals one patch width. Frames are assigned by pose translation, and an anchor is retained only if its complete 128×128 patch lies inside its region. Input and reference points are clipped to that region before filtering or aggregation, preventing processing support and accumulated observations from crossing split boundaries. Supplementary Method S1 identifies the accompanying split and source-frame records.

Within each region, eligible anchors are processed in increasing order of frame index. An anchor is retained if its horizontal center is at least 1 m from every previously retained center. Selection uses geometry only and is fixed before model performance is examined; patches are generated from the selected anchors without pixel-based sliding-window extraction. The resulting split has 129 training, 23 validation, and 27 test patches. The combined sets of input and reference frames contain 1251, 320, and 474 unique frames, respectively, with no shared frame across splits.

Each sparse input uses one anchor scan, retaining native ring IDs 3, 7, 11, ..., 63 before the preprocessing in Section 2.2. All four channels use only these selected returns. The reference uses other full 64-ring scans from the same assigned region and excludes the anchor. Figure 3 shows the physical input channels. Input–reference pairs are generated as follows:

(1) Fix the spatial regions and anchor list. Use the world-aligned 0.20 m grid and anchor-relative height datum in Section 2.2. Construct all four input channels without reference-only returns or semantic labels.

(2) Select same-region reference scans at frame-index offsets −160 to +160 in steps of two, excluding the anchor. At most 160 frames are used. Actual counts are 129–160 for training, 118–158 for validation, and 131–160 for testing.

(3) Apply the sensor region of interest. Use the supplied poses to transform reference returns to world coordinates and clip them to the assigned spatial region. Retain semantic IDs 1, 3, 10, 23, 33, and 34, and project the selected returns onto the world-aligned patch grid using the anchor-relative height datum. These labels select reference surfaces only; the reference branch does not use the input CSF procedure.

(4) Calculate each contributing frame's mean height in each cell. Their median is the candidate reference height. Base validity requires at least six pooled returns from at least three frames, pooled point-height standard deviation ≤ 0.15 m, and frame-mean interquartile range ≤ 0.15 m.

(5) Split reference frames into those before and after the anchor. Each subset must independently meet the same validity criteria, and their median heights must agree within

0.15 m. Final validity requires both the base and temporal tests. A missing or invalid subset invalidates the cell, without interpolation or fallback. The target for a retained cell is the median over the full reference window.

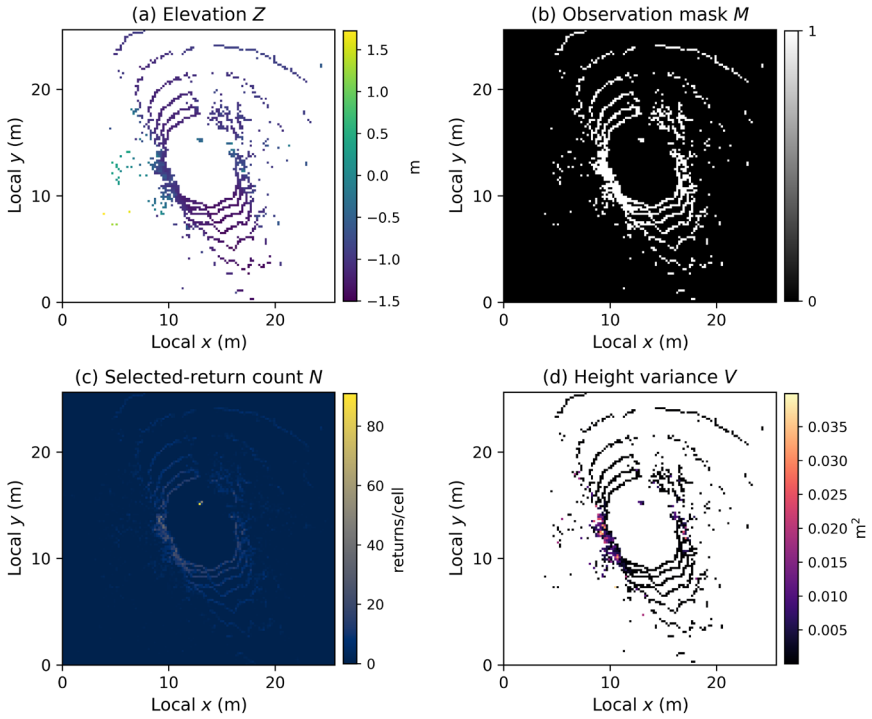

Figure 3. Four physical input channels for sample r5_00000_000320: (**a**) elevation Z (m), (**b**) observation mask M, (**c**) selected-return count N, and (**d**) population height variance V (m^2). White in (**a**,**d**) denotes invalid input cells; $M = 1$ denotes an observed input cell. N is a return count per 0.04 m^2 cell and can be positive where $M = 0$. Coordinates are measured from the patch origin.

The temporal-consistency rule was fixed using training-region diagnostics before the validation and test sets were generated. This screening retains cells with consistent observations across the two temporal subsets.

Let Ω_{ref} denote the set of reference-valid cells. The supervision and reconstruction-evaluation domain is:

$$D = \{\text{i} : \text{M}_\text{i} = 0 \text{ and } \text{M}_{\text{ref,i}} = 1\} \tag{17}$$

An input-missing cell contributes to the loss and reconstruction metrics only when a valid reference exists. Invalid stored zeros are excluded using the masks. Table 2 reports the coverage and distribution of the reconstruction domain across patches; all percentages use the full 16,384-cell patch as their denominator.

Table 2. Reference and reconstruction-domain coverage across the fixed spatial split. The last column reports the minimum, median, and maximum number of eligible cells per patch.

Split	Patches	Reference-Valid (%)	D Mean (%)	D Cells: Min/Median/Max
Training	129	18.75	15.53	1555/2458/4027
Validation	23	18.03	15.99	2168/2561/3233
Test	27	24.13	20.38	2523/3211/5004

The reconstruction domains contain 328,232 training cells, 60,256 validation cells, and 90,157 test cells.

3.2. Models and Training Protocol

Four models are evaluated in a two-by-two comparison of the explicit mask channel and the bottleneck enhancement module. Base U-Net uses $[Z, N, V]$ without enhancement; Mask U-Net uses $[Z, M, N, V]$ without enhancement; Freq U-Net uses $[Z, N, V]$ with enhancement; and MGFE-Net uses $[Z, M, N, V]$ with enhancement. All variants share the encoder–decoder width, spatial split, normalization, loss, and exact observation backfilling. Omitting the mask from the network input does not remove its use in supervision-domain construction or backfilling. This comparison isolates the two tested factors while retaining count and variance channels in all four models.

Each model is trained from scratch with seeds 42, 101, and 202, giving 12 runs. Training uses Python 3.11.14 and PyTorch 2.7.1+cu128 on an NVIDIA GeForce RTX 5090 D v2. The Adam optimizer uses an initial learning rate of 10^{-4}, zero weight decay, first- and second-moment decay rates of 0.9 and 0.999, and an epsilon of 10^{-8}. The batch size is eight, and training runs for a maximum of 200 epochs. CosineAnnealingLR uses a 200-epoch period and a minimum learning rate of zero. No data augmentation is applied. Normalization follows Section 2.2. Supplementary Method S2 provides further implementation and reproducibility details.

Each run retains the checkpoint with the lowest pooled validation RMSE in meters. Training stops after 20 consecutive epochs without a decrease in validation RMSE. Test results are not used for checkpoint selection.

Within each run, reconstruction metrics are pooled over eligible cells or edges, and risk metrics use the denominators defined in Section 3.3. Each of the three runs receives equal weight in the across-seed mean, and dispersion is reported as sample standard deviation (SD).

Efficiency is measured using each model's validation-selected seed-42 checkpoint on the same GPU. A fixed synthetic GPU-resident input has shape $1 \times 4 \times 128 \times 128$; each model selects its designated channels. Inference uses FP32, with automatic mixed precision and TF32 disabled. Forward prediction and completion including backfilling are timed independently using synchronized CUDA events. Each mode has 50 warm-up calls and five groups of 200 timed calls. Mean latency uses 1000 calls; sample SD uses the five group means. Parameter counts are obtained directly, and GMACs include Conv/Linear weight operations only. Loading, point-cloud preprocessing, transfers, and terrain evaluation are excluded. Supplementary Method S4 specifies the timing scope and additional checks.

Two additional checks complement the primary three-seed comparison: a reconstruction-only MGFE-Net trained with seed 42 and evaluated on the fixed test split, and a comparison of the same predictions before and after backfilling on validation samples. Their protocols and results are reported in Supplementary Method S4 and Tables S5 and S6.

3.3. Reconstruction Metrics and Missing-Condition Analysis

All reconstruction errors are evaluated in physical meters after inverse height normalization. For $i \in D$, let $e_i = \hat{Z}_i - Z_{\mathrm{ref},i}$. Root mean square error and mean absolute error are defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i \in D} e_i^{\,2}}{|\mathrm{D}|}}, \quad \mathrm{MAE} = \frac{\sum_{i \in D} |\mathrm{e_i}|}{|\mathrm{D}|} \tag{18}$$

P95 and P99 are percentiles of pooled absolute height errors. All four metrics have units of meters. GE follows the eligible-edge definition in Section 2.4 and has units of m/m: both endpoints are reference-valid and at least one belongs to D. Internal GE uses two missing endpoints; boundary GE uses one missing and one observed endpoint. Crop-border and reference-invalid edges are excluded. The test set has 139,796 internal and 30,901 boundary edges.

RMSE is evaluated across groups defined by the fraction of reference-valid cells with missing input heights in each patch, four-connected hole size within the reconstruction domain, distance to the nearest observed input cell, and input-return density in a clipped 5 × 5 neighborhood. Density is calculated from raw input-return counts, including those in cells classified as missing, and excludes reference returns. Distance is measured to observed input cells rather than to the sensor. All group boundaries are fixed before test inference; missing-fraction boundaries use training quartiles. Supplementary Method S3 and Table S1 define the groups and missing-fraction distribution, and Supplementary Table S2 reports grouped RMSE for all four models.

Terrain evaluation distinguishes feature computability from reference agreement. Slope uses central differences with valid direct neighbors. Step height and roughness use the range and population SD in a complete valid 3 × 3 neighborhood. Each surface has its own validity mask. Coverage is evaluated at reference-valid centers, while feature errors use common feature-valid cells. Exploratory high-risk thresholds are 0.10 m step height, 15° slope, and 0.04 m roughness. Reaching any threshold gives high risk. Reaching 70% of any threshold, without reaching a high threshold, gives medium risk; otherwise, risk is low. Missing features remain unknown. These thresholds are geometric proxies, not calibrated vehicle safety limits.

Within the common all-feature-valid domain, false-low fractions use either reference-high or predicted-low cells as their denominator. Strict false connections are pairs of reference-low components connected through predicted-low cells but disconnected after reference-high cells are removed. A four-neighbor A* check shares reference-derived endpoints across surfaces and allows movement only through predicted-low cells. Availability is the fraction of samples with reference endpoints that yield a predicted path; failed searches contribute no successes. Reference-high crossing fractions use successful-path steps only and are undefined if no such steps exist. Supplementary Method S3 gives the exact event and path rules; Table S4 reports additional feature results and path statuses.

4. Results

4.1. Reconstruction Accuracy and Component Analysis

Table 3 compares the four variants on the fixed spatial holdout, using 90,157 input-missing/reference-valid cells per run. MGFE-Net achieves the lowest mean RMSE, MAE, P95, P99, and overall GE. Relative to Base U-Net, mean RMSE decreases by 6.77%, from 0.1708 to 0.1593 m, and mean GE by 7.95%, from 0.1550 to 0.1427 m/m. MAE decreases by 8.98%. The remaining P99 of 0.4702 m indicates substantial local errors despite the lower averages.

Table 3. Reconstruction and two-factor comparison, mean ± sample SD over seeds 42, 101, and 202. Height errors are in meters; GE is in m/m. All variants use exact observation backfilling. Input-channel and enhancement settings follow Section 3.2.

Model	RMSE (m)	MAE (m)	P95 (m)	P99 (m)	GE (m/m)
Base U-Net	0.1708 ± 0.0053	0.1128 ± 0.0059	0.3174 ± 0.0108	0.4879 ± 0.0129	0.1550 ± 0.0036
Mask U-Net	0.1702 ± 0.0159	0.1096 ± 0.0099	0.3212 ± 0.0536	0.5270 ± 0.0906	0.1481 ± 0.0081
Freq U-Net	0.1697 ± 0.0048	0.1122 ± 0.0060	0.3073 ± 0.0046	0.4926 ± 0.0084	0.1522 ± 0.0004
MGFE-Net	0.1593 ± 0.0056	0.1026 ± 0.0076	0.2865 ± 0.0112	0.4702 ± 0.0013	0.1427 ± 0.0024

The mask channel and enhancement module each give small mean RMSE reductions when added alone; their combination gives the lowest mean error. Mask U-Net has the largest across-seed RMSE SD (0.0159 m). Boundary GE exceeds internal GE for every model. For MGFE-Net, these are 0.3197 ± 0.0148 and 0.1036 ± 0.0044 m/m; for Base U-Net,

they are 0.3759 ± 0.0188 and 0.1062 ± 0.0013 m/m. Supplementary Table S3 reports all individual-seed errors, including internal and boundary GE.

4.2. Qualitative Comparison and Training Behavior

Figure 4 shows three test cases selected before model prediction. The selection criteria are the largest connected hole within the reconstruction domain, the largest within-patch 95th-percentile distance to observed input cells, and the lowest median input density. Together, these criteria identify r5_00000_000320, r5_00000_000509, and r5_00000_000517. The main figures use seed 42. All models produce dense outputs, with differences in the reconstructed terrain between observed cells.

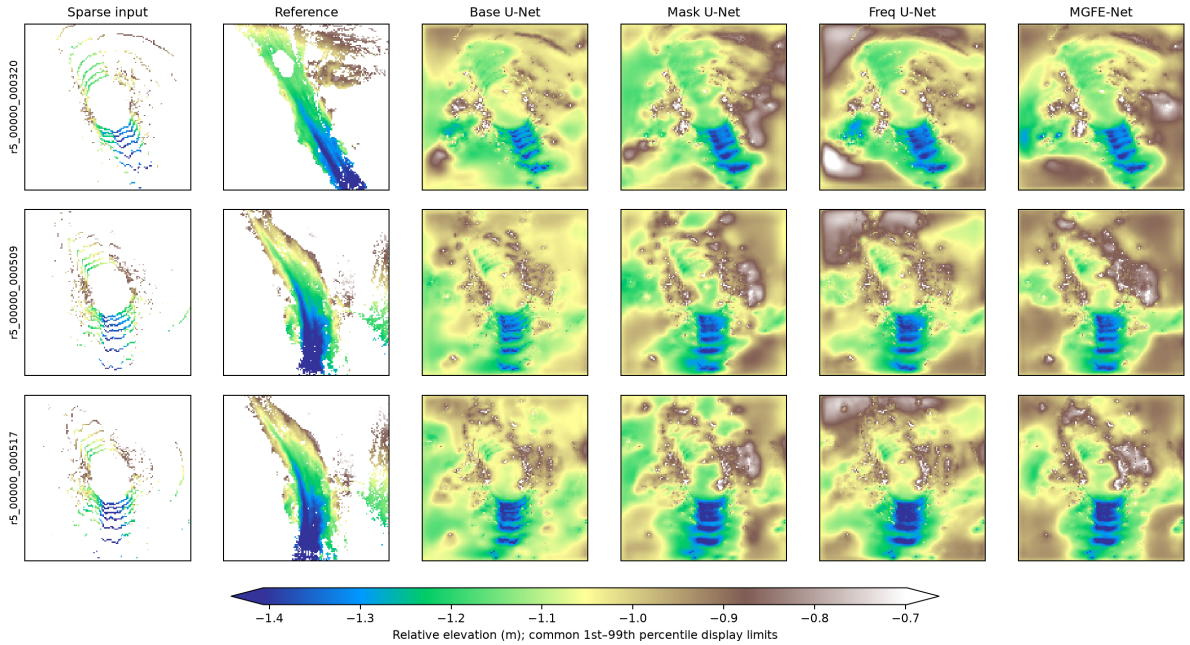

Figure 4. Preselected test cases: sparse input, accumulated reference, and the four seed-42 completions. Each patch spans 25.6 × 25.6 m. Panels share the displayed 1st–99th percentile height limits; saturation changes display only. White denotes unavailable input or reference heights.

Figure 5 shows a rapid initial decrease in validation error, followed by smaller fluctuations. At each epoch, the curves summarize only the seeds that reached that epoch, so later segments have fewer contributors.

4.3. Performance Across Missing Conditions

Table 4 reports all 17 predefined groups for Base U-Net and MGFE-Net, together with eligible-cell counts. MGFE-Net has the lowest mean RMSE among all four variants in 16 groups. For holes of 157–600 cells, Base U-Net achieves an RMSE of 0.0846 m, compared with 0.0863 m for MGFE-Net. Supplementary Table S2 provides the full four-model comparison.

The highest missing-fraction group contains 20 of 27 patches and 79.37% of eligible cells, dominating the aggregate score (Supplementary Table S1). Error increases with distance to observations: MGFE-Net RMSE is 0.3416 m at distances ≥1.6 m and 0.1131 m below 0.4 m. Errors are also larger in lower-density groups. The non-monotonic relationship with hole size shows that area alone does not capture reconstruction difficulty.

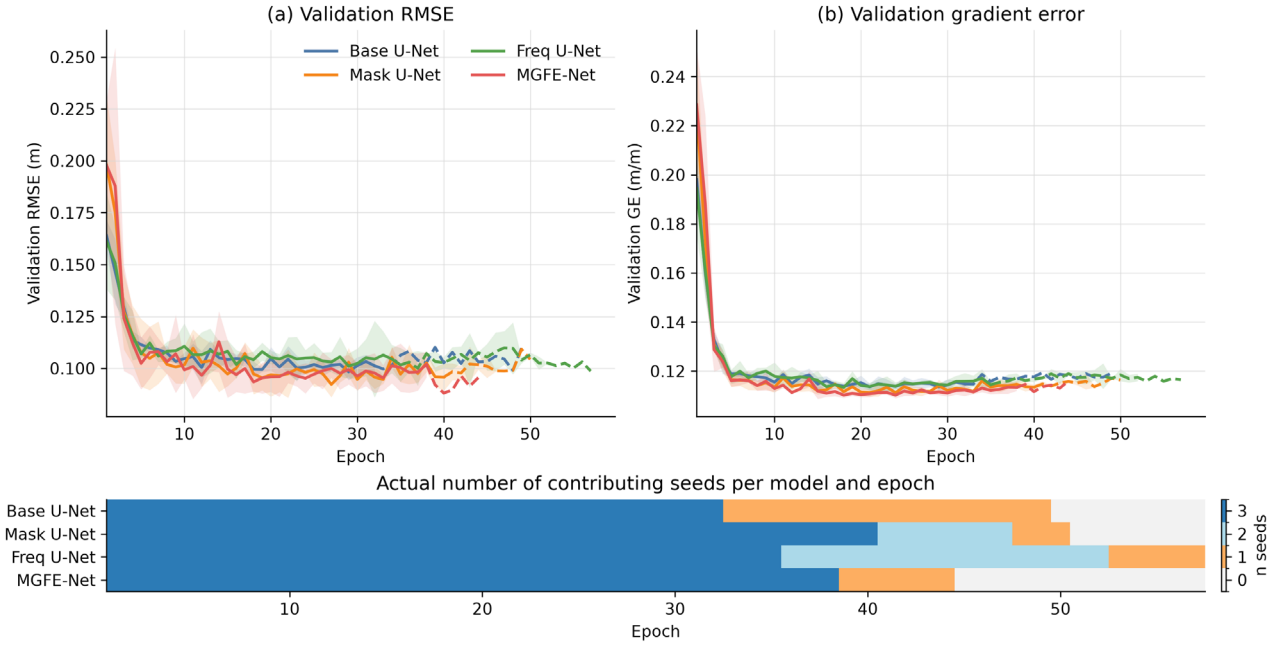

Figure 5. Validation RMSE (m), GE (m/m), and contributing seed counts. Solid lines use $n = 3$ seeds; dashed lines use $n < 3$. Bands are ±1 sample SD and are omitted for $n < 2$; no post-stop values are padded. Selected epochs for seeds 42/101/202 are Base: 12/12/29; Mask: 20/30/27; Freq: 15/37/32; MGFE: 18/24/18.

Table 4. Missing-condition distributions and RMSE (m), mean ± sample SD. Counts are eligible cells per run. Missing fraction is measured within reference-valid cells; labels are rounded, while assignment uses full-precision boundaries. Distance is to an observed input cell, not sensor range.

Factor	Bin	Cells	Base U-Net	MGFE-Net
Missing fraction	<0.817780	2523	0.1320 ± 0.0056	0.1229 ± 0.0032
	[0.817780, 0.826355)	5137	0.1499 ± 0.0092	0.1404 ± 0.0106
	[0.826355, 0.836566)	10,938	0.1970 ± 0.0080	0.1842 ± 0.0069
	≥0.836566	71,559	0.1691 ± 0.0047	0.1576 ± 0.0052
Hole size (cells)	<157	19,227	0.2564 ± 0.0014	0.2455 ± 0.0072
	[157, 601)	8679	0.0846 ± 0.0034	0.0863 ± 0.0013
	[601, 799)	12,880	0.1769 ± 0.0106	0.1572 ± 0.0104
	≥799	49,371	0.1350 ± 0.0085	0.1223 ± 0.0151
Distance (m)	<0.4	32,420	0.1217 ± 0.0023	0.1131 ± 0.0012
	[0.4, 0.8)	30,804	0.1517 ± 0.0060	0.1380 ± 0.0076
	[0.8, 1.6)	21,479	0.1889 ± 0.0094	0.1736 ± 0.0131
	≥1.6	5454	0.3518 ± 0.0030	0.3416 ± 0.0019
Density (returns/m^2)	<1	28,633	0.2085 ± 0.0087	0.1961 ± 0.0111
	[1, 12)	17,381	0.1961 ± 0.0044	0.1822 ± 0.0035
	[12, 36)	23,799	0.1443 ± 0.0037	0.1320 ± 0.0028
	[36, 71)	11,985	0.1085 ± 0.0019	0.1014 ± 0.0017
	≥71	8359	0.0980 ± 0.0013	0.0918 ± 0.0034

4.4. Terrain Features and Exploratory Risk

The proportion of reference-valid cell centers at which all three terrain features can be computed increases from 0.44% for sparse inputs to 98.94% for each completed map. The corresponding proportion is 59.61% for the reference surface. All dense variants achieve identical coverage because they produce finite elevations on the same grid. Table 5 reports feature errors and risk outcomes, with additional coverage and feature statistics in Supplementary Table S4.

Table 5. Terrain-feature errors and exploratory risk results, reported as mean ± sample SD over three seeds. Feature errors use common feature-valid cells. Risk and path denominators follow Section 3.3 and Supplementary Method S3.

Metric	Base	Mask	Freq	MGFE
Slope RMSE (°)	9.206 ± 0.168	8.975 ± 0.389	8.970 ± 0.045	8.796 ± 0.036
Step-height RMSE (m)	0.1259 ± 0.0010	0.1253 ± 0.0064	0.1244 ± 0.0008	0.1240 ± 0.0039
Roughness RMSE (m)	0.0424 ± 0.0004	0.0423 ± 0.0023	0.0421 ± 0.0003	0.0417 ± 0.0017
Reference-high predicted low (%)	40.51 ± 1.16	38.74 ± 2.47	40.14 ± 1.53	40.10 ± 1.89
Predicted-low that is reference-high (%)	11.77 ± 0.43	11.06 ± 0.49	11.41 ± 0.27	11.06 ± 0.23
Strict false connection (%)	2.74 ± 0.15	2.67 ± 0.50	2.50 ± 0.30	2.84 ± 0.68
Path availability (%)	2.47 ± 2.14	0.00 ± 0.00	1.23 ± 2.14	6.17 ± 7.71

MGFE-Net achieves the lowest mean feature RMSE: 8.796° for slope, 0.1240 m for step height, and 0.0417 m for roughness. However, 40.10 ± 1.89% of reference-high cells are classified as low risk, and 11.06 ± 0.23% of predicted-low cells are reference-high. Its strict false-connection rate is 2.84 ± 0.68%, compared with 2.74 ± 0.15% for Base U-Net. Figure 6 shows the reference-conditioned comparisons, and Supplementary Figure S1a–c presents risk maps across the three seeds. The lower reconstruction errors therefore do not correspond to a lower mean strict false-connection rate.

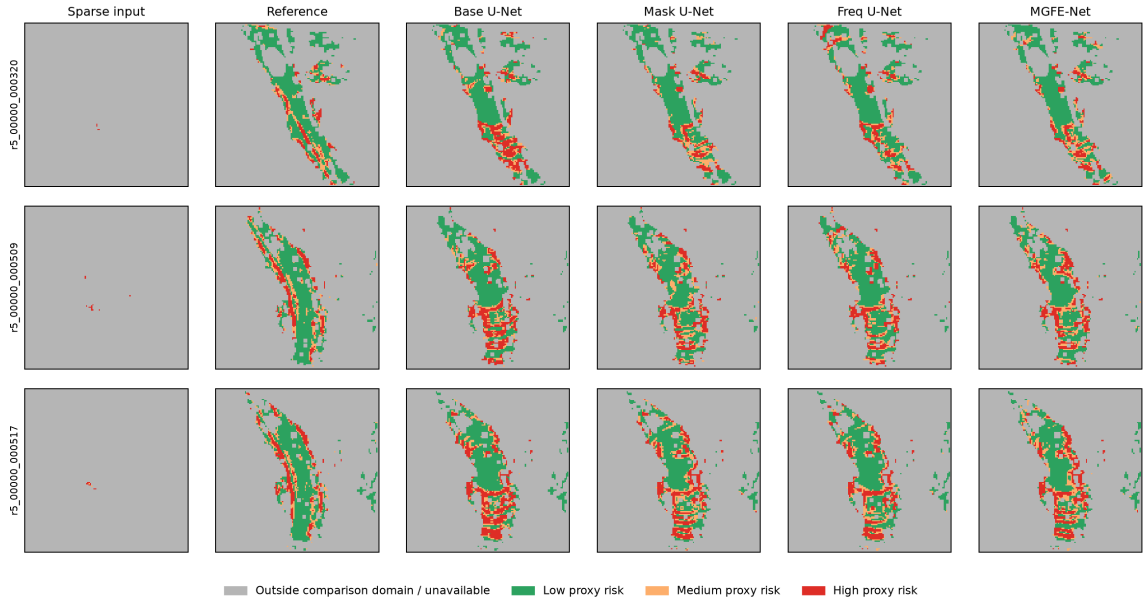

Figure 6. Exploratory risk classes for the same cases and seed as Figure 4. Panels use the common reference all-feature-valid display domain. Gray denotes unavailable risk classifications or cells outside that domain.

MGFE-Net finds a path in 5 of 81 sample–seed attempts, compared with 2 of 81 for Base U-Net. Across all four models, 8 attempts succeed, 135 have no predicted path, and 181 have endpoints that are not predicted low risk. Sparse-input paths are unavailable under the same endpoint criteria. Supplementary Table S4 provides the complete path-search outcomes.

4.5. Computational Cost and Supplementary Checks

As summarized in Table 6, MGFE-Net has 16,807,873 trainable parameters and 7.932 Conv/Linear GMACs. Mean forward and completion latencies are 4.225 and 4.188 ms, respectively, both higher than the corresponding Base U-Net means. The modes were timed independently; their small reversal does not indicate that backfilling accelerates inference. Timing excludes preprocessing, transfers, and terrain evaluation.

Table 6. Trainable parameters, Conv/Linear GMACs, and GPU-resident inference latency on an NVIDIA GeForce RTX 5090 D v2 (FP32, batch size 1, 128 × 128). Latency is mean ± sample SD of five timing-group means; completion includes backfilling. Measurement scope follows Section 3.2 and Supplementary Method S4.

Model	Parameters	GMACs	Forward (ms)	Completion (ms)
Base U-Net	13,395,329	7.730	3.798 ± 0.006	3.809 ± 0.012
Mask U-Net	13,395,905	7.740	3.821 ± 0.016	3.830 ± 0.011
Freq U-Net	16,807,297	7.923	4.183 ± 0.011	4.195 ± 0.015
MGFE-Net	16,807,873	7.932	4.225 ± 0.045	4.188 ± 0.025

In the post hoc seed-42 comparison, removing gradient supervision increases test RMSE from 0.1608 to 0.1741 m and boundary GE from 0.3027 to 0.5165 m/m (Supplementary Table S5). In the validation-only backfilling check, missing-cell RMSE stays at 0.0890 m and the maximum observed-cell difference from the input falls from 1.8276 m to zero. Boundary GE increases from 0.1048 to 0.1840 m/m (Supplementary Table S6). These checks distinguish observation retention from reference agreement and are separate from the three-seed component comparison.

5. Discussion

MGFE-Net achieves the lowest mean reconstruction errors among the four variants. The mask channel makes the distinction between observed and missing cells available during feature extraction, while the enhancement module combines pooled context with residual details at the bottleneck. Together, these components provide information about measurement locations and terrain variation at different spatial scales.

Exact backfilling preserves observed input heights when completion supplements existing measurements. Agreement with the accumulated reference can nevertheless deteriorate at boundaries: in the validation check, missing-cell predictions stay unchanged while boundary GE increases. Likewise, a preserved center height can acquire different slope, step height, or roughness when its neighbors are completed. Observation retention and neighborhood accuracy are distinct properties.

The downstream results distinguish feature availability from risk accuracy. Dense predictions permit more local feature calculations, but false-low classifications and strict false connections remain. Smoothing an unseen discontinuity can create an apparent low-risk bridge despite a lower average height error. A planner should retain the observation mask to distinguish measured and inferred terrain. Future work could calibrate height-error bounds, propagate them through terrain-feature operators, and use conservative costs or an unknown state when those bounds cross a risk threshold.

6. Limitations

Evaluation is restricted to one spatial split of RELLIS-3D sequence 00000. Although source frames are disjoint across splits, patches within each region remain spatially correlated. The reported SDs quantify variation across training seeds on this split. Statistical

significance and variation across independent scenes were not assessed. Generalization to other scenes, datasets, and sensor configurations requires further evaluation.

Multi-frame references approximate a selected terrain surface and remain susceptible to pose errors, mixed surfaces, and viewpoint-dependent sampling. Temporal-consistency screening may preferentially retain stable, well-supported surfaces. On average, reference-valid cells account for 24.13% of the cells in each test patch, while the reconstruction domain covers 20.38%. The reported errors therefore characterize only the portion of terrain with valid reference support.

Comparisons are restricted to four U-Net variants. The count and variance channels remain fixed across variants, so their individual contributions are unresolved. The gradient-loss comparison uses one seed, and the backfilling analysis compares validation outputs from the same forward predictions. Downstream evaluation uses geometric thresholds and grid-based paths without vehicle-specific dynamics. End-to-end onboard performance and traversability accuracy require validation under operating conditions.

7. Conclusions

This study presented MGFE-Net for completing sparse LiDAR-derived elevation maps using elevation, observation mask, selected-return count, and height variance. The network combines a U-Net-like encoder–decoder with pooling-based multiscale enhancement and exact observation backfilling. Evaluation on a documented spatial holdout of RELLIS-3D sequence 00000 used 129 training, 23 validation, and 27 test patches, with three training seeds per model. MGFE-Net achieved an RMSE of 0.1593 ± 0.0056 m and a GE of 0.1427 ± 0.0024 m/m, reducing the Base U-Net means by 6.77% and 7.95%, respectively.

Completion increased terrain-feature computability, but improved reconstruction averages did not produce a lower mean strict false-connection rate. The evidence therefore supports reconstruction improvement under the tested sensor subsampling and reference construction, while leaving traversability and planning safety unverified. Independent-scene evaluation, additional baselines, calibrated predictive uncertainty, and onboard validation are needed to assess broader use.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/vehicles8110289/s1, Supplementary Methods S1–S4, Tables S1–S6, and Figure S1a–c are provided in the accompanying supplementary document. Supporting code and numerical records are included in the associated archive.

Author Contributions: Conceptualization, L.Z., Y.Z. (Yuchen Zhang) and F.G.; methodology, L.Z. and Y.Z. (Yuchen Zhang); software, L.Z. and Y.Z. (Yuchen Zhang); validation, Y.Z. (Yuchen Zhang), Y.Z. (Yuhang Zhang) and C.L.; formal analysis, L.Z. and Y.Z. (Yuchen Zhang); investigation, Y.Z. (Yuhang Zhang), C.L. and X.L.; resources, F.G. and C.L.; data curation, L.Z., Y.Z. (Yuchen Zhang) and Y.Z. (Yuhang Zhang); writing—original draft preparation, L.Z. and Y.Z. (Yuchen Zhang); writing—review and editing, F.G., C.L. and X.L.; visualization, L.Z. and Y.Z. (Yuchen Zhang); supervision, F.G.; project administration, F.G.; funding acquisition, F.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Inner Mongolia, grant numbers 2024MS05049 and 2026SHZR3322; and the Inner Mongolia Autonomous Region Science and Technology Plan Project, grant number 2026YFHH0231.

Data Availability Statement: The raw data are publicly available in the RELLIS-3D dataset [20]. This study uses sequence 00000. Processed split manifests, evaluation masks, implementation code, and experimental records are provided in the accompanying Supplementary Files.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Definition
LiDAR	Light detection and ranging
RMSE	Root mean square error
MAE	Mean absolute error
GE	Gradient error
MSE	Mean squared error
SD	Standard deviation
IQR	Interquartile range
CSF	Cloth simulation filtering
BN	Batch normalization
ReLU	Rectified linear unit
FP32	32-bit floating-point arithmetic
GMACs	Billions of multiply–accumulate operations

References

1. Papadakis, P. Terrain traversability analysis methods for unmanned ground vehicles: A survey. *Eng. Appl. Artif. Intell.* **2013**, *26*, 1373–1385. [CrossRef]
2. Borges, P.V.K.; Peynot, T.; Liang, S.; Arain, B.; Wildie, M.; Minareci, M.G.; Lichman, S.; Samvedi, G.; Sa, I.; Hudson, N.; et al. A survey on terrain traversability analysis for autonomous ground vehicles: Methods, sensors, and challenges. *Field Robot.* **2022**, *2*, 1567–1627. [CrossRef]
3. Fankhauser, P.; Bloesch, M.; Gehring, C.; Hutter, M.; Siegwart, R. Robot-centric elevation mapping with uncertainty estimates. In *Mobile Service Robotics*; World Scientific: Singapore, 2014; pp. 433–440. [CrossRef]
4. Fankhauser, P.; Hutter, M. A universal grid map library: Implementation and use case for rough terrain navigation. In *Robot Operating System (ROS): The Complete Reference*; Koubaa, A., Ed.; Springer: Cham, Switzerland, 2016; Volume 1, pp. 99–120. [CrossRef]
5. Yang, B.; Zhang, Q.; Geng, R.; Wang, L.; Liu, M. Real-time neural dense elevation mapping for urban terrain with uncertainty estimations. *IEEE Robot. Autom. Lett.* **2023**, *8*, 696–703. [CrossRef]
6. Stölzle, M.; Miki, T.; Gerdes, L.; Azkarate, M.; Hutter, M. Reconstructing occluded elevation information in terrain maps with self-supervised learning. *IEEE Robot. Autom. Lett.* **2022**, *7*, 1697–1704. [CrossRef]
7. Shan, T.; Englot, B.; Meyers, D.; Wang, W.; Ratti, C.; Rus, D. LIO-SAM: Tightly-coupled LiDAR inertial odometry via smoothing and mapping. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 25–29 October 2020; pp. 5135–5142. [CrossRef]
8. Uhrig, J.; Schneider, N.; Schneider, L.; Franke, U.; Brox, T.; Geiger, A. Sparsity invariant CNNs. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 11–20. [CrossRef]
9. Ma, F.; Karaman, S. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 4796–4803. [CrossRef]
10. Eldesokey, A.; Felsberg, M.; Khan, F.S. Confidence propagation through CNNs for guided sparse depth regression. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2423–2436. [CrossRef] [PubMed]
11. Xu, Y.; Zhu, X.; Shi, J.; Zhang, G.; Bao, H.; Li, H. Depth completion from sparse LiDAR data with depth-normal constraints. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2811–2820. [CrossRef]
12. Khan, M.A.U.; Nazir, D.; Pagani, A.; Mokayed, H.; Liwicki, M.; Stricker, D.; Afzal, M.Z. A comprehensive survey of depth completion approaches. *Sensors* **2022**, *22*, 6969. [CrossRef] [PubMed]
13. Hu, J.; Bao, C.; Ozay, M.; Fan, C.; Gao, Q.; Liu, H.; Lam, T.L. Deep depth completion from extremely sparse data: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 8244–8264. [CrossRef] [PubMed]
14. Tang, J.; Tian, F.-P.; An, B.; Li, J.; Tan, P. Bilateral propagation network for depth completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 17–21 June 2024; pp. 9763–9772. [CrossRef]

15. Yan, Z.; Wang, Z.; Wang, K.; Li, J.; Yang, J. Completion as enhancement: A degradation-aware selective image guided network for depth completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 11–15 June 2025; pp. 26943–26953. [CrossRef]
16. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]
17. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.-C.; Tao, A.; Catanzaro, B. Image inpainting for irregular holes using partial convolutions. In *Computer Vision—ECCV 2018*; Springer: Cham, Switzerland, 2018; Volume 11215, pp. 89–105. [CrossRef]
18. Suvorov, R.; Logacheva, E.; Mashikhin, A.; Remizova, A.; Ashukha, A.; Silvestrov, A.; Kong, N.; Goka, H.; Park, K.; Lempitsky, V. Resolution-robust large mask inpainting with Fourier convolutions. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2022; pp. 3172–3182. [CrossRef]
19. Jiang, L.; Dai, B.; Wu, W.; Loy, C.C. Focal frequency loss for image reconstruction and synthesis. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 13899–13909. [CrossRef]
20. Jiang, P.; Osteen, P.R.; Wigness, M.; Saripalli, S. RELLIS-3D dataset: Data, benchmarks and analysis. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 1110–1116. [CrossRef]
21. Zhang, W.; Qi, J.; Wan, P.; Wang, H.; Xie, D.; Wang, X.; Yan, G. An easy-to-use airborne LiDAR data filtering method based on cloth simulation. *Remote Sens.* **2016**, *8*, 501. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.