

Article

Super-Resolving Digital Terrain Models Using a Modified RCAN

Mohamed Helmy^{1,2,3,*} , Emanuele Mandanici¹ , Luca Vittuari¹  and Gabriele Bitelli¹ 

¹ Faculty of Engineering, Department of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, 40136 Bologna, Italy; emanuele.mandanici@unibo.it (E.M.); luca.vittuari@unibo.it (L.V.); gabriele.bitelli@unibo.it (G.B.)

² Faculty of Engineering, Geodesy and Geomatics Division, Sapienza University of Rome, 00184 Rome, Italy

³ Faculty of Engineering, Department of Civil Engineering, University of Benha, Benha 13512, Egypt

* Correspondence: mohamedmohsen.helmy@unibo.it; Tel.: +39-351-3001718

Highlights

What are the main findings?

- A modified Residual Channel Attention Network (RCAN) effectively super-resolves 10 m DTMs to 1 m spatial resolution using elevation-aware loss functions.
- The balanced loss configuration ($\alpha = 0.5$, $\gamma = 0.5$) achieved the lowest MAE (0.83 m) and RMSE (1.14–1.15 m), maintaining near-zero bias across diverse terrain types.

What are the implications of the main findings?

- The proposed deep learning framework enables accurate high-resolution DTM generation in regions lacking LiDAR coverage.
- This approach supports operational applications in geomorphology, hydrology, and landscape monitoring by providing detailed and reliable elevation data from coarse inputs.

Abstract

High-resolution Digital Terrain Models (DTMs) are essential for precise terrain analysis, yet their production remains constrained by the high cost and limited coverage of LiDAR surveys. This study introduces a deep learning framework based on a modified Residual Channel Attention Network (RCAN) to super-resolve 10 m DTMs to 1 m resolution. The model was trained and validated on a 568 km² LiDAR-derived dataset using custom elevation-aware loss functions that integrate elevation accuracy (L1), slope gradients, and multi-scale structural components to preserve terrain realism and vertical precision. Performance was evaluated across 257 independent test tiles representing flat, hilly, and mountainous terrains. A balanced loss configuration ($\alpha = 0.5$, $\gamma = 0.5$) achieved the best results, yielding Mean Absolute Error (MAE) as low as 0.83 m and Root Mean Square Error (RMSE) of 1.14–1.15 m, with near-zero bias (−0.04 m). Errors increased moderately in mountainous areas (MAE = 1.29–1.41 m, RMSE = 1.84 m), confirming the greater difficulty of rugged terrain. Overall, the approach demonstrates strong potential for operational applications in geomorphology, hydrology, and landscape monitoring, offering an effective solution for high-resolution DTM generation where LiDAR data are unavailable.

Keywords: DTMs; neural networks; RCAN; super-resolution; accuracy assessment



Academic Editors: Takashi Oguchi and Tomaž Podobnikar

Received: 11 November 2025

Revised: 15 December 2025

Accepted: 19 December 2025

Published: 21 December 2025

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Digital Terrain Models (DTMs) are fundamental tools for analyzing and modeling Earth surface processes, supporting applications in hydrology, geomorphology, natural hazard assessment, and environmental planning [1–3]. They provide raster-based, geo-referenced representations of the Earth's surface elevation, capturing topographic variability essential for surface water modeling, landscape evolution, and risk analysis. High-resolution DTMs, typically derived from airborne LiDAR or UAV photogrammetry, offer detailed depictions of terrain morphology but are limited by high acquisition costs, complex logistics, and restricted spatial coverage, particularly in remote or densely vegetated areas [4–6].

In contrast, global DTMs such as the Shuttle Radar Topography Mission (SRTM), ALOS World 3D, and Copernicus GLO-30 provide near-global coverage at spatial resolutions of 30 m or more, enabling regional to global-scale studies [7,8]. However, these coarser models often fail to represent fine-scale topographic details, especially in mountainous or urban environments, reducing their suitability for high-precision modeling tasks. This trade-off between spatial resolution and coverage has motivated the development of super-resolution (SR) methods that aim to computationally enhance the resolution of coarse DTMs, providing a cost-effective alternative for terrain refinement [9–12].

Super-resolution techniques, originally developed in the image-processing field, reconstruct high-resolution images from one or more low-resolution inputs [13]. When applied to elevation data, the challenge lies in recovering fine-scale geomorphological structures from sparse inputs. Traditional SR approaches can be broadly classified into interpolation-based, reconstruction-based, and learning-based methods [14,15]. Interpolation approaches such as bilinear, bicubic, Kriging, or Inverse Distance Weighted (IDW) are simple and computationally efficient but tend to over smooth terrain features, failing to preserve ridges, cliffs, or slopes [16,17]. Reconstruction-based methods use gradient or edge constraints to improve surface detail, yet their effectiveness declines under high upscaling factors or in heterogeneous terrain [18,19].

Learning-based methods, especially deep learning (DL) approaches, have recently shown great promise by learning non-linear relationships between low-resolution and high-resolution elevation data. Early frameworks relied on manifold learning [20], sparse coding [21], or patch-based dictionary matching [10], but these were computationally expensive. The introduction of convolutional neural networks (CNNs) enabled end-to-end training from paired datasets, and models such as Super-Resolution Convolutional Neural Network (SRCNN), Enhanced Deep Residual Networks (EDSR), Super-Resolution Generative Adversarial Network (SRGAN), Enhanced SRGAN (ESRGAN), and Residual Channel Attention Network (RCAN), originally designed for natural image enhancement, have been successfully adapted for DTM super-resolution [22–27]. Among these, the Residual Channel Attention Network (RCAN) is particularly effective due to its ability to recover high-frequency terrain details using residual blocks and channel-wise attention [28].

Despite these advances, applying image-based models directly to elevation data presents unique challenges. DTMs represent continuous surfaces with geometric properties such as slope, aspect, and curvature, which are not typically considered in models trained on natural images. As a result, standard CNN-based methods may produce elevation inconsistencies, loss of geomorphological structure, or artifacts affecting drainage networks and valley boundaries [29,30]. Furthermore, deep learning models often generalize poorly across different terrain types and rarely incorporate topographic constraints or uncertainty quantification [31,32].

To address these limitations, recent studies have integrated terrain-specific descriptors such as slope, curvature, and roughness into model architectures or loss functions [26,33].

These terrain-aware strategies reinforce physical consistency and improve the realism of reconstructed surfaces. Multi-component loss functions combining pixel-wise accuracy (L1, L2), perceptual similarity (SSIM), and gradient-based terrain consistency have further enhanced both numerical precision and structural coherence [25,34]. Hybrid frameworks, such as detrending-based deep learning (DTDLD), have also been proposed to separate large-scale elevation trends from high-frequency residuals, allowing more effective learning of fine-scale patterns [35].

Beyond deep learning, probabilistic and geostatistical approaches have treated DTM super-resolution as a non-unique reconstruction problem, modeling multiple plausible fine-resolution surfaces consistent with the same coarse data [9,36]. Methods based on variograms [37,38] or Multiple-Point Statistics (MPS) using training images [39–41] can reproduce spatial structures realistically but are computationally demanding and rely on suitable training data [42,43].

Despite substantial methodological progress, generating accurate high-resolution DTMs remains a challenge, especially in topographically complex regions such as high mountain Asia [44], where global DTMs like SRTM or ASTER GDEM fail to capture sharp terrain discontinuities. Preserving key landform features such as ridgelines, drainage channels, and valley floors is critical for hydrological and environmental modeling [45,46]. Moreover, performance must be assessed not only through error metrics such as RMSE or MAE but also in terms of geomorphological realism and through the consistency of derived terrain parameters [47,48].

The present study proposes a deep learning-based super-resolution framework for DTM enhancement using the RCAN model, optimized through terrain-aware loss functions that incorporate domain-specific elevation and slope information. The research aims to (1) identify and evaluate the most effective loss function for terrain super-resolution and (2) fine-tune the corresponding weights to balance elevation accuracy and structural preservation. Using a 568 km² LiDAR-derived dataset covering diverse terrain types, the model is trained to generate 1 m resolution DTMs from 10 m inputs. Model performance is evaluated through statistical and structural metrics across flat, hilly, and mountainous areas. The results demonstrate that combining deep learning with geomorphological informed loss functions provides a practical and scalable approach for producing high-resolution DTMs from widely available coarse datasets.

2. Materials and Methods

2.1. Overview of the Workflow

The proposed workflow is structured into two main phases. In the first phase, several custom loss functions were developed and assessed to determine which configuration best improves DTM reconstruction quality. In the second phase, the selected loss function was fine-tuned by adjusting the relative balance between its components to further optimize model performance. The backbone of the super-resolution framework is the Residual Channel Attention Network (RCAN), an advanced neural architecture designed to upscale spatial data while preserving fine-scale structural details. For this study, RCAN was specifically adapted for terrain data by incorporating loss components sensitive to elevation-dependent features such as slope gradients.

2.2. Dataset Description

The dataset used in this study consists of 568 high-resolution Digital Terrain Model (DTM) tiles freely provided by the Italian Ministry of the Environment and Energy Security (MASE) (<https://sim.mase.gov.it/portalediaccesso/mappe/#/viewer/new> (accessed on 14 November 2024)) (Figure 1). Each tile covers a 1 km × 1 km region and was originally

generated from airborne LiDAR acquisitions at a spatial resolution of 1 m. The tiles span a wide range of morphological settings, including flat alluvial plains, rolling hills, and steep mountainous terrain, ensuring that the experimental framework captures diverse topographic conditions. All DTMs were downloaded in GeoTIFF format with complete geospatial metadata, allowing consistent spatial alignment and seamless integration into the processing and modeling pipeline.

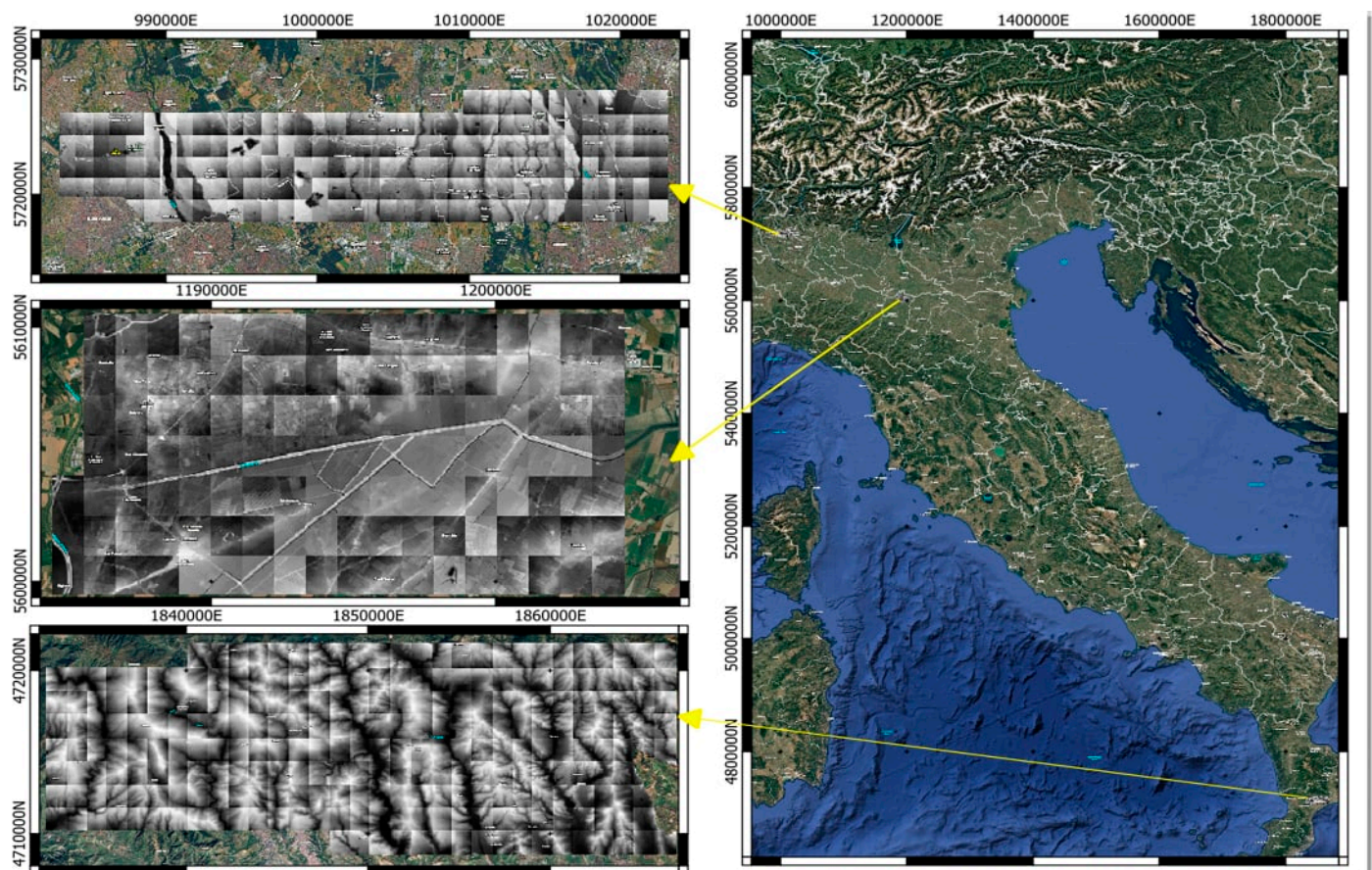


Figure 1. Visualization of the dataset used in the super-resolution model training validation and testing.

2.3. Data Preprocessing

Prior to model training, each high-resolution DTM tile (1 m; 1000×1000 pixels) was downsampled to 10 m (100×100 pixels) using average aggregation to simulate realistic coarse-resolution DTM data while preserving the underlying terrain structure. The original georeferencing metadata was retained and propagated to both resolutions to ensure consistent spatial reference during evaluation. The resulting paired low- and high-resolution tiles were then divided into training, validation, and testing subsets so that model performance could be assessed on unseen terrain.

2.4. RCAN Model Architecture

The network used in this study builds on the Residual Channel Attention Network (RCAN), with several adaptations made to accommodate the characteristics of single-band Digital Terrain Models (DTMs) and the required $10\times$ spatial upscaling (Figure 2). The original RCAN organizes its feature extraction into multiple Residual Groups (RGs), each containing several Residual Channel Attention Blocks (RCABs). For the present work, this structure was simplified to a single sequence of ten RCABs. This adjustment reduces

computational load while still allowing the network to extract detailed spatial information from elevation data.

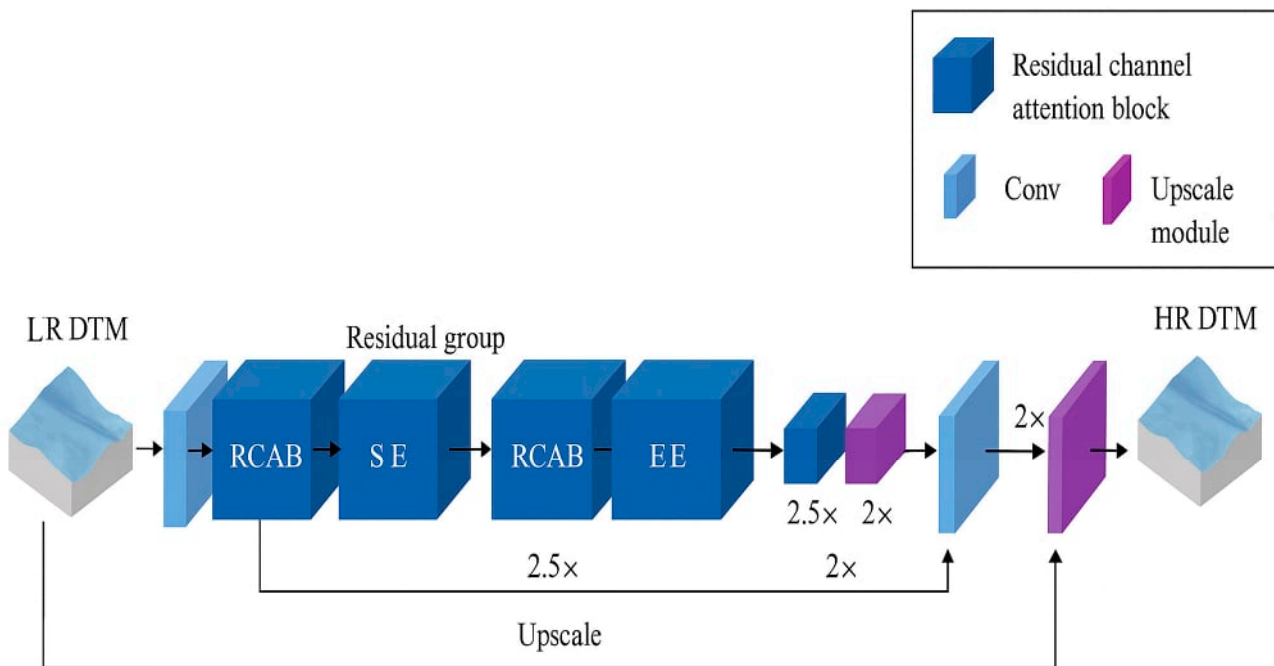


Figure 2. Modified RCAN-Based DTM Super-Resolution workflow.

Each RCAB retains the standard components of RCAN: two convolution layers with ReLU activation and a Squeeze-and-Excitation channel-attention module. All convolution layers use reflect padding rather than the zero padding adopted in the original implementation. Reflect padding helps reduce boundary artifacts that can otherwise appear along tile edges and propagate through the network—an important consideration when working with geospatial datasets.

A more substantial modification concerns the upsampling module. The classical RCAN uses pixel-shuffle layers designed for integer upscaling factors (e.g., $2\times$ or $4\times$). Since the task here requires a non-integer global scaling factor of $10\times$ (from 10 m to 1 m), a progressive interpolation strategy was introduced. The final upsampling unit consists of three stages: an initial $2.5\times$ bilinear interpolation, followed by two $2\times$ bilinear interpolation steps. Each stage is paired with a convolution layer and a LeakyReLU activation to refine intermediate representations. This approach avoids the checkerboard artifacts often associated with sub-pixel convolution when applied to non-integer scaling and provides more stable optimization during training.

The architecture was also adapted to the single-channel structure of DTMs by replacing the RGB-based input and output layers of the original model with one-channel convolution kernels. This ensures that the network handles elevation values consistently and preserves the metric nature of the data.

2.5. Custom Loss Functions

Let $\hat{Y} \in R^{B \times 1 \times H \times W}$ be the predicted DTM batch and $Y \in R^{B \times 1 \times H \times W}$ the ground truth. Index pixels by $p = (i, j)$. Let $N = H \times W$ be the per-sample pixel count. Define forward finite differences $(\nabla_x X)_{i,j} = X_{i,j+1} - X_{i,j}$, $(\nabla_y X)_{i,j} = X_{i+1,j} - X_{i,j}$,

$$i,j = 1 - X_{i,j}, (\nabla_y X)_{i,j} = X_{i+1,j} - X_{i,j}.$$

2.5.1. LF1

$$L_1(\hat{Y}, Y) = \frac{1}{BN} \sum_{b=1}^B \sum_{i=1}^H \sum_{j=1}^W |\hat{Y}_{b,j,i} - Y_{b,j,i}|$$

$$L_2(\hat{Y}, Y) = \frac{1}{BN} \sum_{b=1}^B \sum_{i=1}^H \sum_{j=1}^W (\hat{Y}_{b,j,i} - Y_{b,j,i})^2$$

$$L_{slope}(\hat{Y}, Y) = \frac{1}{B} \sum_{b=1}^B \left(\frac{1}{N_x} \sum_{i,j} \left| (\nabla_x \hat{Y})_{b,i,j} - (\nabla_x Y)_{b,i,j} \right| + \frac{1}{N_y} \sum_{i,j} \left| (\nabla_y \hat{Y})_{b,i,j} - (\nabla_y Y)_{b,i,j} \right| \right)$$

$$LF1 = \alpha \times L_1 + \beta \times L_2 + \gamma \times L_{slope}$$

2.5.2. LF2

$$L_1(\hat{Y}, Y) = \frac{1}{BN} \sum_b \sum_{i,j} |\hat{Y}_{b,j,i} - Y_{b,j,i}|$$

$$h_{mid,b} = \frac{1}{2} \left(\min_{i,j} Y_{b,i,j} + \max_{i,j} Y_{b,i,j} \right)$$

$$\omega_{b,i,j} = e^{-\frac{|Y_{b,j,i} - h_{mid,b}|}{100}}$$

$$L_{elev}(\hat{Y}, Y) = \frac{1}{BN} \sum_b \sum_{i,j} \omega_{b,i,j} |\hat{Y}_{b,j,i} - Y_{b,j,i}|$$

$$LF2 = \alpha \times L_1 + \beta \times L_{elev}$$

2.5.3. LF3

Gaussian blur kernel size = 5, $\sigma = 1.0$, padding = reflect, number of pyramid levels $L = 3$, and downsampling via average pooling by factor 2 at each level.

$$L_1(\hat{Y}, Y) = \frac{1}{BN} \sum_b \sum_{i,j} |\hat{Y}_{b,j,i} - Y_{b,j,i}|$$

Define $B(\cdot)$ as Gaussian blur and $D(\cdot)$ (avg pool downsample) Laplacian residual at level k : $\ell_{lap}^{(k)}(X) = X^{(k)} - B(X^{(k)})$, $X^{(0)} = X$, $X^{(k+1)} = D(B(X^{(k)}))$

$$L_{lap}(\hat{Y}, Y) = \sum_{k=0}^{L-1} \frac{1}{BN^{(k)}} \sum_b \sum_p \left| \ell_{lap}^{(k)}(\hat{Y})_{b,p} - \ell_{lap}^{(k)}(Y)_{b,p} \right|$$

$$L_{grad}(\hat{Y}, Y) = \frac{1}{B} \sum_b \left(\frac{1}{N_x} \sum_{i,j} \left| (\nabla_x \hat{Y})_{b,i,j} - (\nabla_x Y)_{b,i,j} \right| + \frac{1}{N_y} \sum_{i,j} \left| (\nabla_y \hat{Y})_{b,i,j} - (\nabla_y Y)_{b,i,j} \right| \right)$$

$$LF3 = \alpha \times L_1 + \beta \times L_{lap} + \gamma \times L_{grad}$$

2.5.4. LF4

$$L_1(\hat{Y}, Y) = \frac{1}{BN} \sum_b \sum_{i,j} |\hat{Y}_{b,j,i} - Y_{b,j,i}|$$

$$L_{slope}(\hat{Y}, Y) = \frac{1}{B} \sum_{b=1}^B \left(\frac{1}{N_x} \sum_{i,j} \left| (\nabla_x \hat{Y})_{b,i,j} - (\nabla_x Y)_{b,i,j} \right| + \frac{1}{N_y} \sum_{i,j} \left| (\nabla_y \hat{Y})_{b,i,j} - (\nabla_y Y)_{b,i,j} \right| \right)$$

$$LF4 = \alpha \times L_1 + \gamma \times L_{slope}$$

To adapt RCAN to the physical characteristics of terrain data, four loss formulations were designed and tested. Each loss integrates pixel-wise error terms with additional components intended to guide the model toward reproducing local slopes, multiscale structure, or elevation-dependent variations.

LF1 combines L1 and L2 errors with a slope-based term derived from horizontal and vertical finite differences. This formulation encourages both absolute accuracy and consistency in local terrain gradients.

LF2 pairs the L1 loss with an elevation-weighted term. The weighting factor assigns higher importance to elevations further from the mean level of each tile, encouraging the network to reduce bias in areas where height variations are more pronounced.

LF3 incorporates a Laplacian-pyramid representation, allowing errors to be evaluated at multiple spatial scales. The multiscale residuals emphasize terrain discontinuities and break lines, while an additional gradient term enforces local slope consistency.

LF4 combines the L1 loss with the slope-based gradient penalty. This simpler formulation targets elevation accuracy while explicitly constraining local gradients, making it well suited for preserving terrain geometry.

Each loss function was tested by training RCAN for 100 epochs using the same training and validation sets and the same $2.5\times \rightarrow 2\times \rightarrow 2\times$ progressive upsampling module. Model selection relied on validation RMSE and SSIM. Among the four candidates, LF4 provided the most stable and accurate results, and was therefore examined further through a small grid search over the weighting parameters (α , γ), with the final configuration selected using the same validation-based procedure. These four configurations (LF1–LF4) constitute an ablation study designed to isolate the contribution of each loss component. Because all models share identical architecture and upsampling, the observed performance differences arise solely from the loss terms, allowing a direct assessment of elevation fidelity, slope consistency, and elevation-weighted components.

2.6. Training Setup and Implementation

All experiments were conducted using the PyTorch 1.0.0 framework on Google Colab (V3) equipped with a Tesla T4 GPU. The learning rate was fixed at 5×10^{-6} , the batch size was set to 4, and random seeds were initialized for reproducibility. Model checkpoints were saved periodically, and the final model was selected based on the lowest validation RMSE across epochs.

2.7. Evaluation Metrics

Model performance was assessed on a per-tile basis using standard statistical and perceptual metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). Visual assessments were also performed through signed error maps and residual histograms to identify spatial patterns of error, particularly in areas with steep slopes or complex morphology.

2.8. Use of Generative AI

Generative Artificial Intelligence (GenAI) tools, specifically ChatGPT 3.5 (OpenAI, San Francisco, CA, USA), were used to assist in the language refinement and formatting of this manuscript. The AI was not used for generating data, designing experiments, analyzing results, or interpreting findings. All scientific content, methods, and conclusions were entirely developed and validated by the authors.

3. Results

3.1. Visual Assessment of Terrain Reconstruction

Figure 3 presents analytical hill shade visualizations for two representative test tiles, comparing the reference LiDAR-derived DTM (1 m), the super-resolved output (1 m), and the original input DTM (10 m). The hill-shaded images highlight clear differences in terrain representation across scales. Compared to the 10 m input, the super-resolved DTMs recover finer geomorphological features, including drainage lines, ridges, and subtle slope transitions, which are largely absent or heavily smoothed in the low-resolution data. While the super-resolved outputs do not fully reproduce all high frequency details present in the reference DTMs, they show a marked improvement in spatial continuity and terrain structure, visually aligning more closely with the reference surfaces.

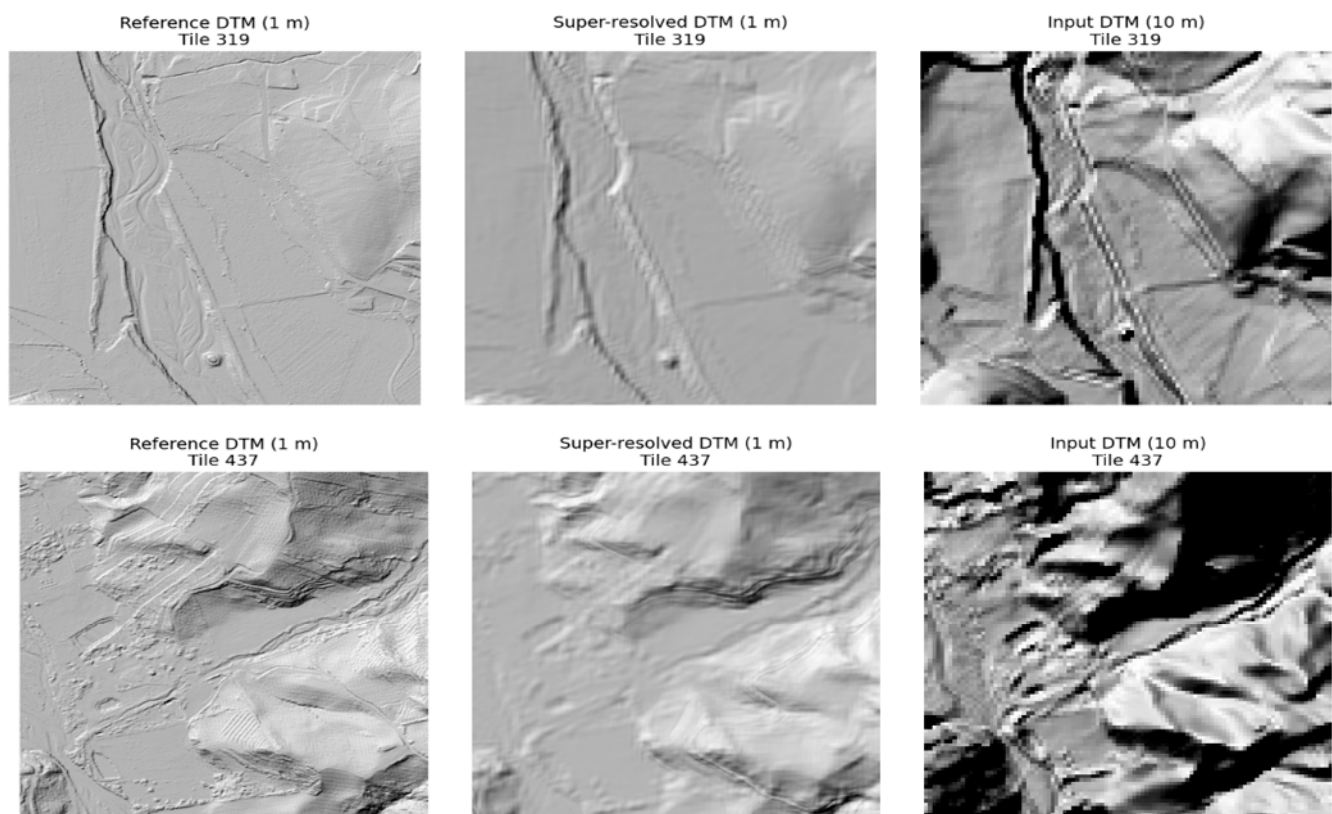


Figure 3. Visual comparison of hill shade representations for selected representative test tiles, using the LF4 loss function.

3.2. Quantitative Evaluation of Loss Functions

To assess the effectiveness of the proposed terrain-aware loss formulations, four distinct loss functions (LF1–LF4) were implemented and tested within the RCAN-based DTM super-resolution framework. Each configuration was trained independently for 100 epochs under identical conditions, including model architecture, dataset, and training parameters, to ensure a consistent basis for comparison. The optimal checkpoint for each configuration was selected according to the minimum validation loss.

Model performance was quantitatively assessed using the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). The averaged results across all test tiles are summarized in Table 1. All loss functions were evaluated using the same progressive $2.5\times \rightarrow 2\times \rightarrow 2\times$ upsampling module, meaning that the architectural contribution to error is constant across configurations. As a result, the RMSE values of the well-behaved loss functions fall within a

relatively narrow range, reflecting the smoothing effect imposed by the interpolation-based upsampling and the corresponding error floor it introduces for recovering high-frequency terrain detail.

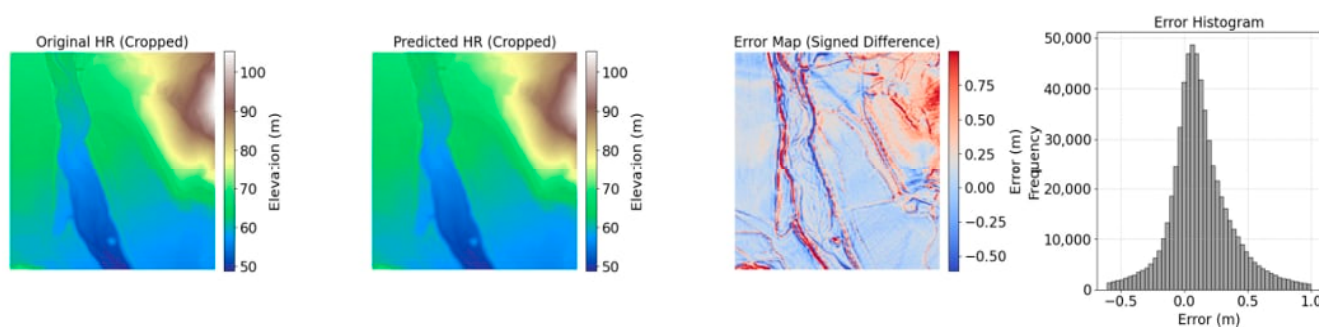
Table 1. Quantitative results of different loss functions using RCAN Super-Resolution technique.

Loss Function	LF1		LF2		LF3		LF4	
	Validation	Testing	Validation	Testing	Validation	Testing	Validation	Testing
RMSE ($\mu \pm \sigma$) (M)	0.28	1.90 ± 0.54	2.82	36.84 ± 13.99	0.27	1.66 ± 0.49	0.28	1.64 ± 0.49
MAE (M)	—	1.45 ± 0.43	—	24.14 ± 9.44	—	1.23 ± 0.39	—	1.21 ± 0.38
Mean PSNR(dB)	59.43	48.01 ± 3.02	40.34	22.59 ± 3.08	59.74	49.24 ± 3.14	59.47	49.30 ± 3.49
Mean SSIM (M)	0.94	0.99 ± 0.01	0.90	0.96 ± 0.01	0.94	0.99 ± 0.01	0.94	0.99 ± 0.01

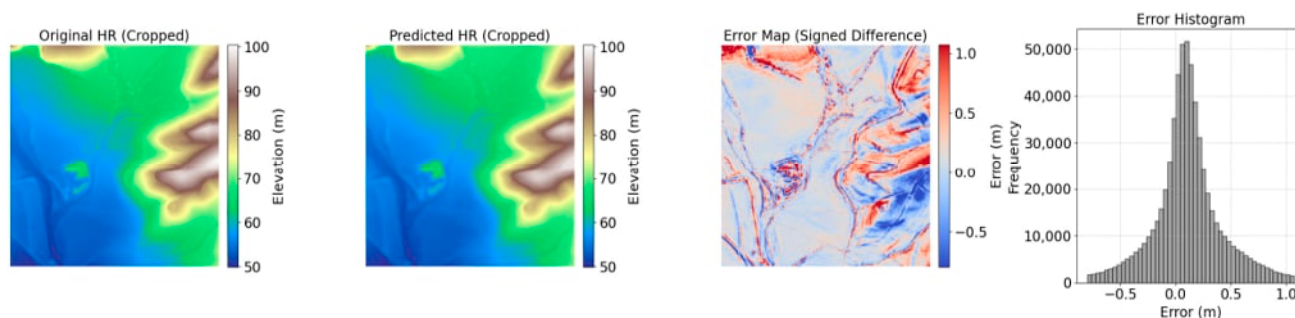
Within this shared architectural context, LF4—combining an L1 elevation term with a slope-consistency component—achieved the best overall performance. It obtained the lowest test RMSE (1.647 ± 0.490) and MAE (1.217 ± 0.382), along with a high SSIM (0.992 ± 0.004) and the highest PSNR (49.304 ± 3.493 dB), indicating strong generalization and improved preservation of both elevation fidelity and terrain structure (Figure 4).

In contrast, LF2, which applies an elevation-weighted L1 penalty relative to deviation from the mean, produced by far the weakest results, with a test RMSE of 36.843 ± 13.991 and an MAE of 24.140 ± 9.442 . This catastrophic drop in accuracy reflects a fundamental limitation of the weighting strategy: the exponential elevation-based weights heavily down-weight high-relief regions, causing the loss signal to collapse precisely where terrain variability is highest. As a result, the model receives almost no gradient information in slopes, ridges, and sharp discontinuities, leading to unstable optimization and highly inconsistent reconstruction behavior (Figure 5).

The distribution of tile-wise statistics for all tested loss functions is illustrated in Figure 6, highlighting that three of the loss formulations exhibit similar performance ranges, while LF4 demonstrates the most accurate and stable results across the dataset.

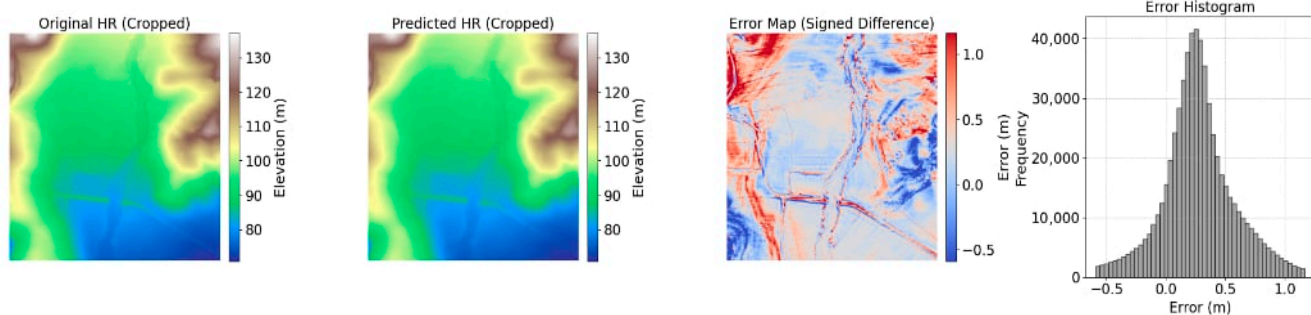


319.tiff—RMSE: 0.378, MAE: 0.242, PSNR: 48.900 dB, SSIM: 0.988

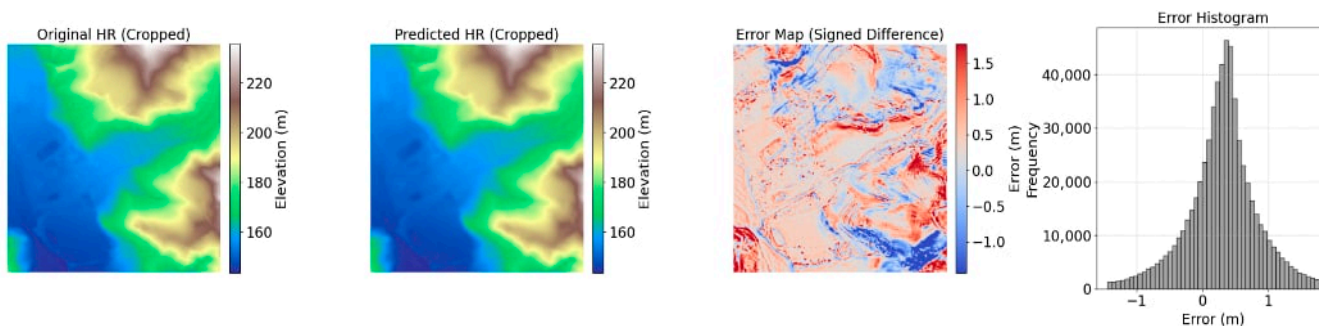


355.tiff—RMSE: 0.421, MAE: 0.279, PSNR: 47.562 dB, SSIM: 0.985

Figure 4. Cont.

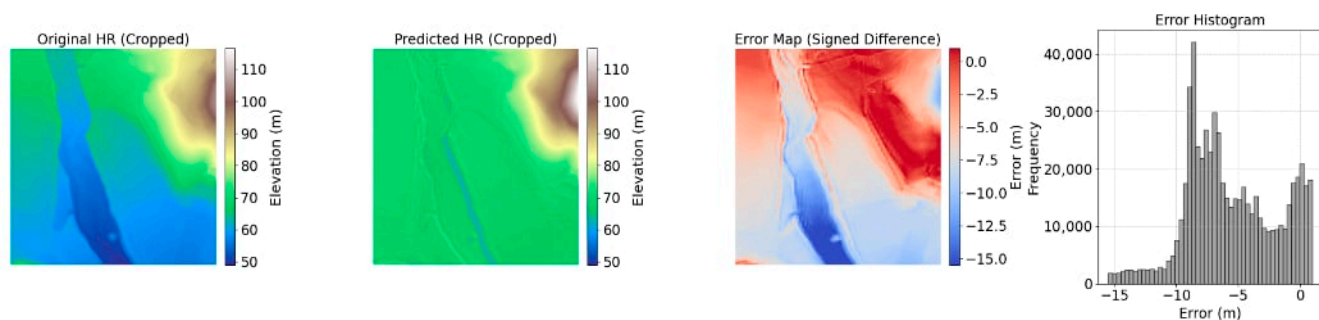


386.tiff—RMSE: 0.479, MAE: 0.364, PSNR: 49.141 dB, SSIM: 0.993

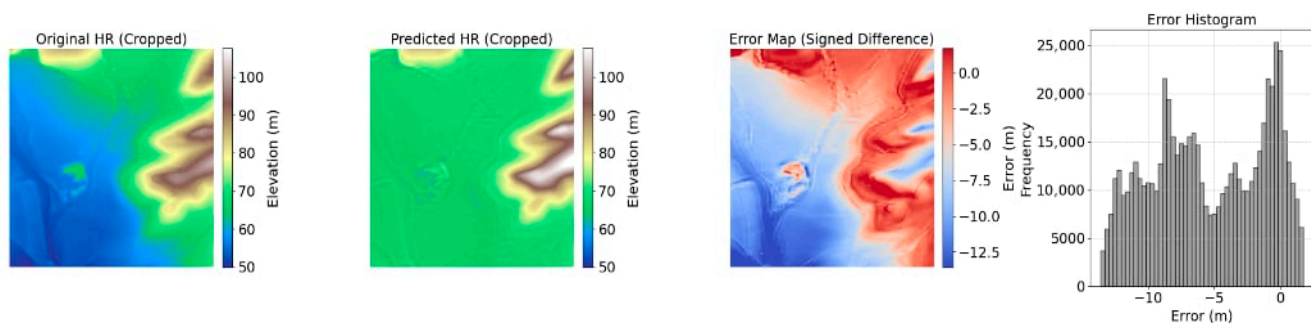


437.tiff—RMSE: 0.756, MAE: 0.557, PSNR: 49.883 dB, SSIM: 0.989

Figure 4. Comparison of the generated super-resolution DEMs with their ground truth for some representative example tiles, using LF4 loss function.

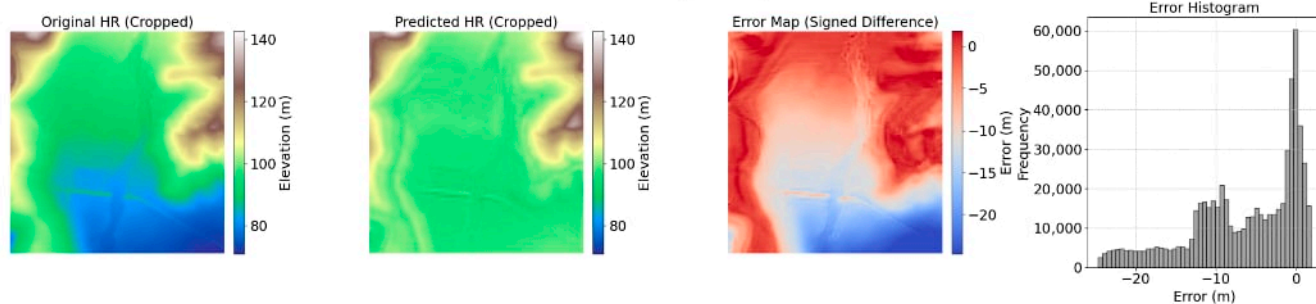


319.tiff—RMSE: 7.023, MAE: 5.865, PSNR: 23.515 dB, SSIM: 0.968

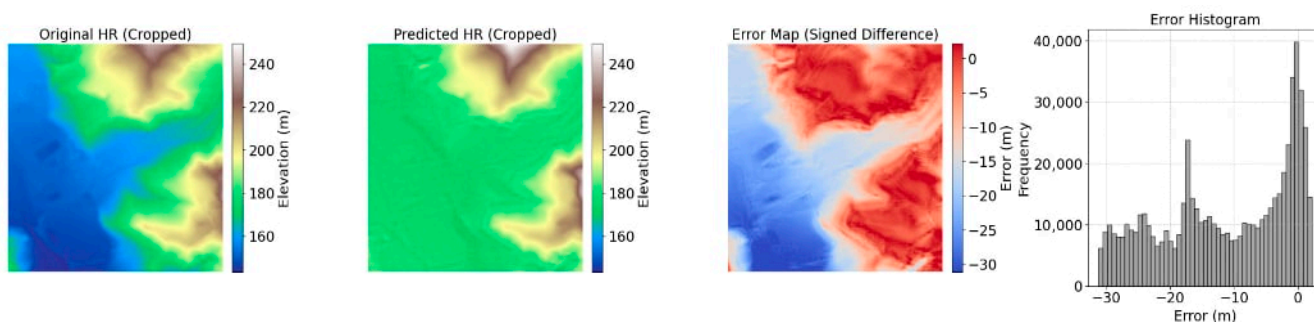


355.tiff—RMSE: 7.084, MAE: 5.686, PSNR: 23.039 dB, SSIM: 0.962

Figure 5. Cont.



386.tif—RMSE: 9.944, MAE: 7.065, PSNR: 22.800 dB, SSIM: 0.976



437.tif—RMSE: 15.495, MAE: 11.752, PSNR: 23.646 dB, SSIM: 0.968

Figure 5. Visualization of some representative generated Super-resolution samples with the ground truth tiles using LF2 loss function.

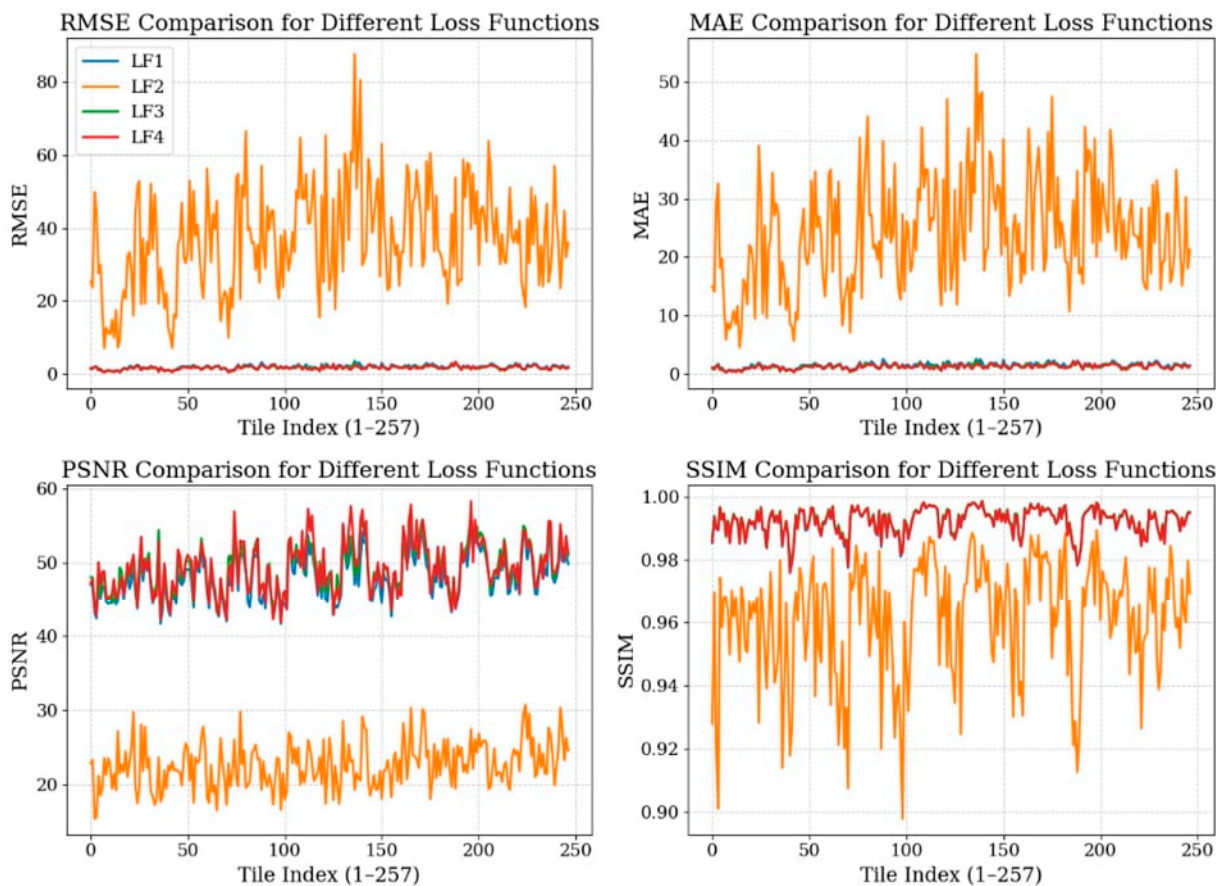


Figure 6. Performance of the four loss functions using tile-wise averaged metrics RMSE, MAE, PSNR, and SSIM.

3.3. Loss Weight Tuning Analysis

To assess the influence of loss weighting on super-resolution performance, three configurations of the combined loss function were tested: $(\alpha = 0.8, \gamma = 0.2)$, $(\alpha = 0.5, \gamma = 0.5)$, and $(\alpha = 0.2, \gamma = 0.8)$, where α represents the weight of the elevation-based term and γ the slope-based term. The objective was to evaluate how emphasizing terrain smoothness versus gradient detail affects model accuracy and generalization.

Visual comparison of the three settings (Figures 7–9) showed minimal perceptible differences across sample tiles, confirming that quantitative assessment provides a more reliable evaluation. As summarized in Table 2, the balanced configuration $(\alpha = 0.5, \gamma = 0.5)$ achieved the lowest RMSE on both validation (0.28) and testing datasets (1.62 ± 0.50) , alongside high PSNR $(49.47 \pm 3.50 \text{ dB})$ and SSIM (0.99 ± 0.01) . This configuration provided the best overall trade-off between elevation accuracy and terrain structure preservation.

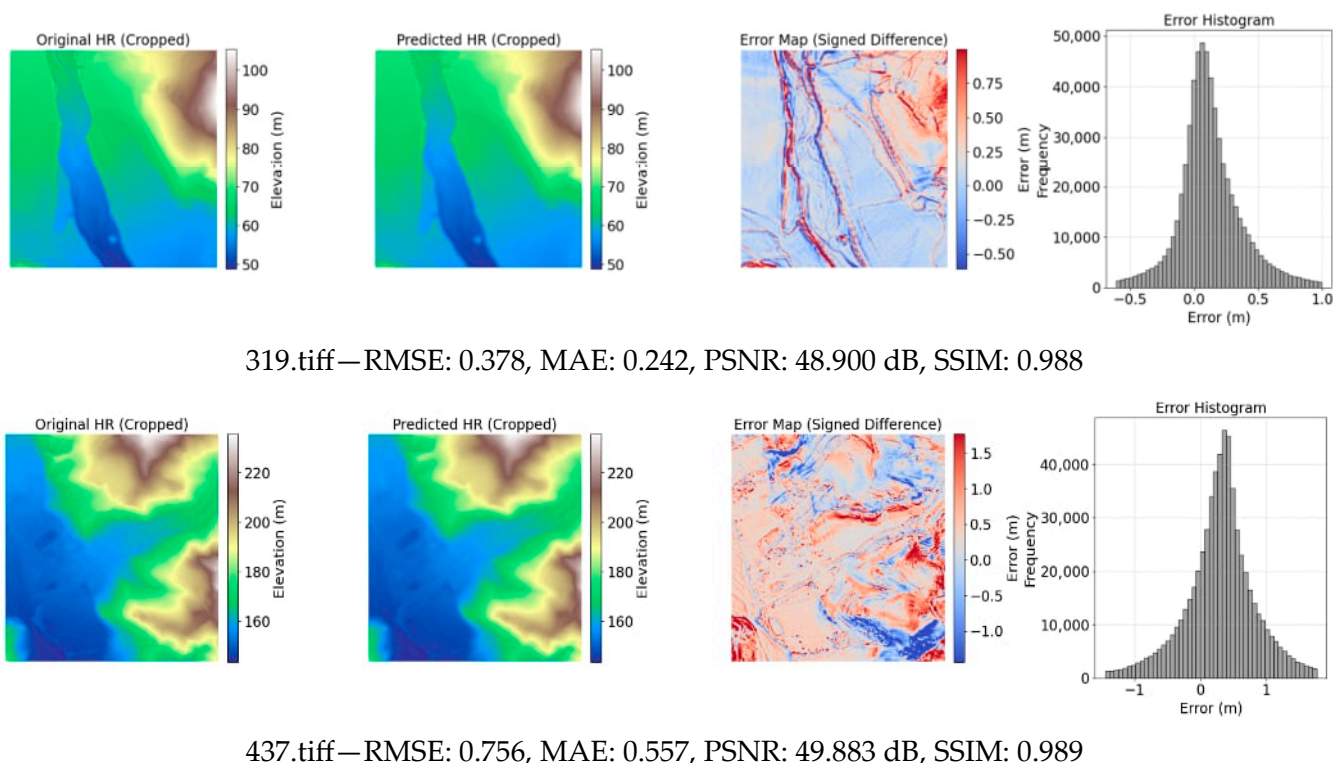
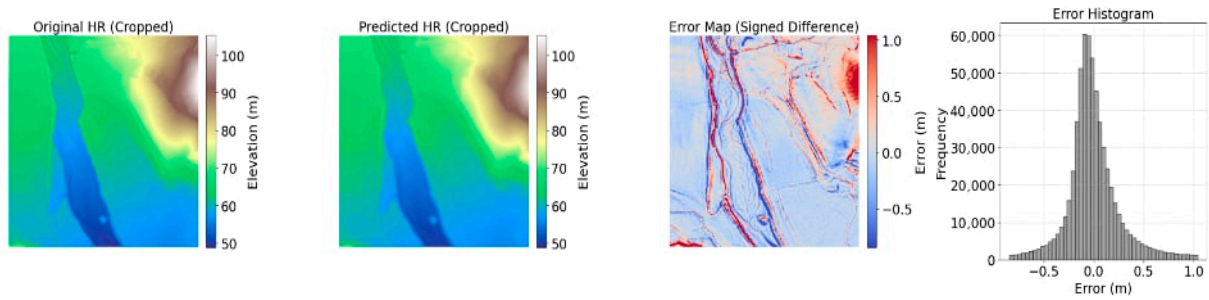


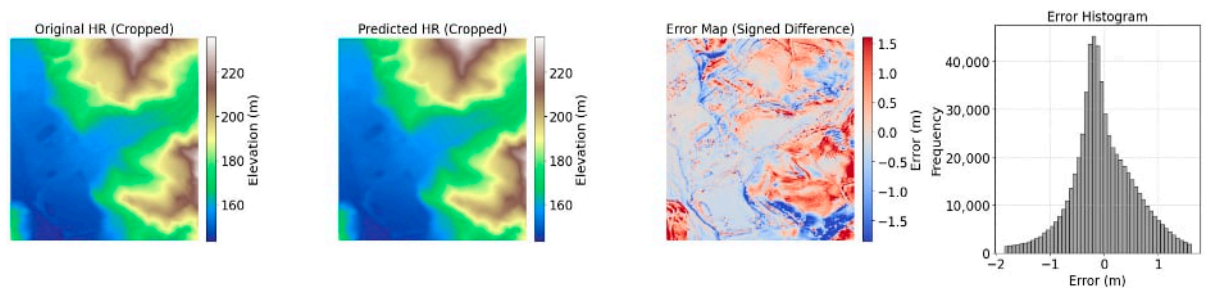
Figure 7. Visualization of the representative generated Super-resolution samples with the ground truth tiles using fine tuning $(\alpha = 0.8, \gamma = 0.2)$ of LF4 loss function.

Table 2. Summarizing of statistical results of the three fine tuning combination of LF4 loss function.

Loss Function	$(\alpha = 0.8, \gamma = 0.2)$		$(\alpha = 0.5, \gamma = 0.5)$		$(\alpha = 0.2, \gamma = 0.8)$	
	Validation	Testing	Validation	Testing	Validation	Testing
RMSE ($\mu \pm \sigma$) (M)	0.28	1.64 ± 0.49	0.28	1.62 ± 0.50	0.28	1.73 ± 0.52
MAE (M)	-	1.21 ± 0.38	-	1.18 ± 0.39	-	1.32 ± 0.42
Mean PSNR (dB)	59.47	49.30 ± 3.49	59.52	49.47 ± 3.50	59.44	48.88 ± 3.10
Mean SSIM (M)	0.94	0.99 ± 0.01	0.94	0.99 ± 0.01	0.94	0.99 ± 0.01

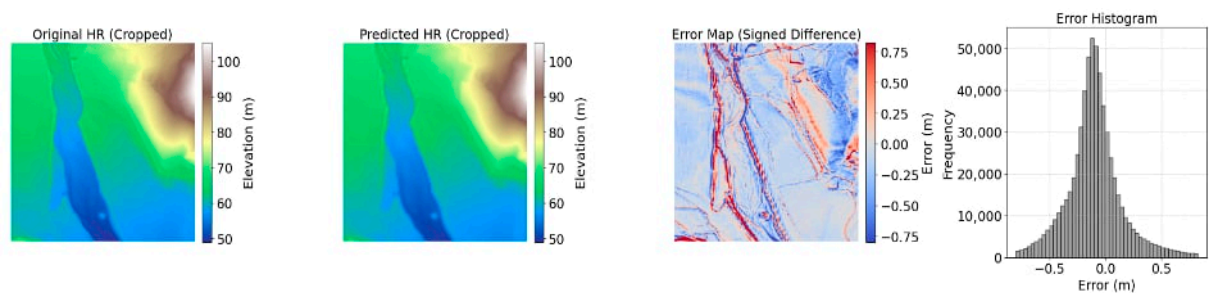


319.tiff—RMSE: 0.391, MAE: 0.232, PSNR: 48.607 dB, SSIM: 0.987

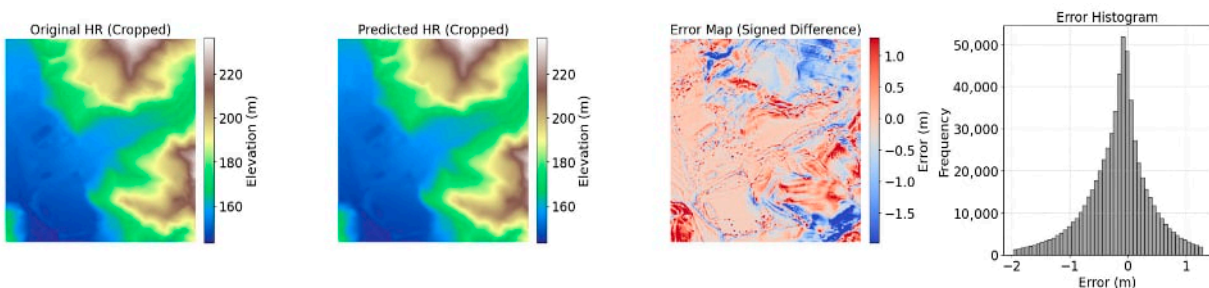


437.tiff—RMSE: 0.763, MAE: 0.531, PSNR: 49.805 dB, SSIM: 0.989

Figure 8. Visualization of the representative generated Super-resolution samples with the ground truth tiles using fine tuning ($\alpha = 0.5$, $\gamma = 0.5$) of LF4 loss function.



319.tiff—RMSE: 0.389, MAE: 0.247, PSNR: 48.644 dB, SSIM: 0.987



437.tiff—RMSE: 0.733, MAE: 0.491, PSNR: 50.151 dB, SSIM: 0.989

Figure 9. Visualization of the representative generated Super-resolution samples with the ground truth tiles using fine tuning ($\alpha = 0.2$, $\gamma = 0.8$) of LF4 loss function.

Increasing the weight of the slope term ($\alpha = 0.2$, $\gamma = 0.8$) led to a higher RMSE (1.73 ± 0.52) and MAE (1.32 ± 0.42), indicating reduced accuracy in absolute elevation, particularly in flat regions. Nonetheless, the SSIM remained consistently high (~ 0.99), reflecting that the overall terrain morphology was well preserved. Conversely, favoring the elevation term ($\alpha = 0.8$, $\gamma = 0.2$) slightly improved RMSE over the slope-heavy variant but did not outperform the balanced configuration.

Per-tile statistics of the test dataset (Figure 10) reveal a steady decline in accuracy as the slope component weight increases. This trend indicates that while slope information enhances structural realism, excessive weighting can compromise elevation precision. Overall, the balanced weighting ($\alpha = 0.5, \gamma = 0.5$) demonstrated the most accurate and consistent results across terrain types.

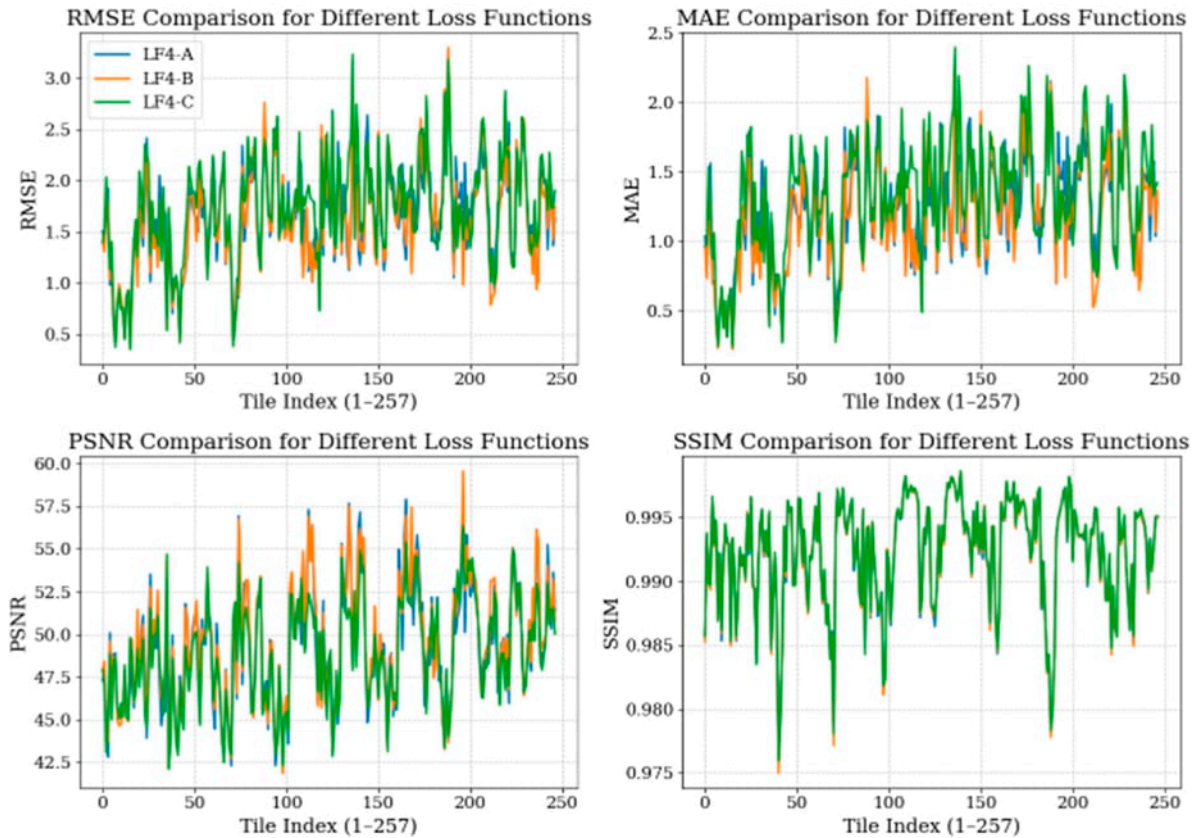


Figure 10. Performance of the three-fine tuning of LF4 (A) ($\alpha = 0.8, \gamma = 0.2$), (B) ($\alpha = 0.5, \gamma = 0.5$), and (C) ($\alpha = 0.2, \gamma = 0.8$) using tile-wise averaged metrics RMSE, MAE, PSNR, and SSIM.

3.4. Terrain-Specific Performance

3.4.1. Loss Functions

To assess how terrain morphology influences super-resolution accuracy, a stratified evaluation was conducted across three terrain classes—flat, hilly, and mountainous. Each test tile was categorized using slope-based masks derived from the ground truth DTM, and performance metrics were computed for each class. The evaluated metrics included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Bias (mean signed error). This stratified approach enabled quantification of model performance under increasing topographic complexity.

The results, illustrated in Figures 11–13 and summarized in Table 3, show that MAE values increase with terrain ruggedness. The LF4 model achieved the lowest MAE overall, with approximately 0.85 m in flat and hilly regions and 1.32 m in mountainous areas. LF3 followed closely with similar values—0.90 m (flat), 0.91 m (hilly), and 1.33 m (mountainous)—demonstrating robust generalization. LF1 produced slightly higher MAE (1.21–1.52 m), while LF2 performed substantially worse, particularly in flat terrain, where MAE reached 30.14 m.

RMSE results confirmed these trends. LF4 yielded the lowest RMSE in flat terrain (1.15 m) and remained under 2 m in mountainous areas (1.75 m). LF3 showed comparable

performance with RMSE values up to 1.76 m, while LF1 ranged between 1.62 m and 1.99 m. LF2 again showed extreme errors, with RMSE peaking at 40.85 m in flat terrain.

Table 3. Slope-based zonal statistics assessment for different loss functions.

Model	Terrain	Mean MAE (m)	Mean RMSE (m)	Mean Bias (m)
LF1	Flat	1.27	1.62	+0.82
	Hilly	1.21	1.57	+0.78
	Mountainous	1.52	1.99	+0.77
LF2	Flat	30.14	40.85	+29.76
	Hilly	26.36	38.50	+25.90
	Mountainous	23.57	35.94	+22.68
LF3	Flat	0.90	1.16	+0.32
	Hilly	0.91	1.19	+0.34
	Mountainous	1.33	1.77	+0.26
LF4	Flat	0.85	1.15	−0.47
	Hilly	0.86	1.17	−0.43
	Mountainous	1.32	1.76	−0.41

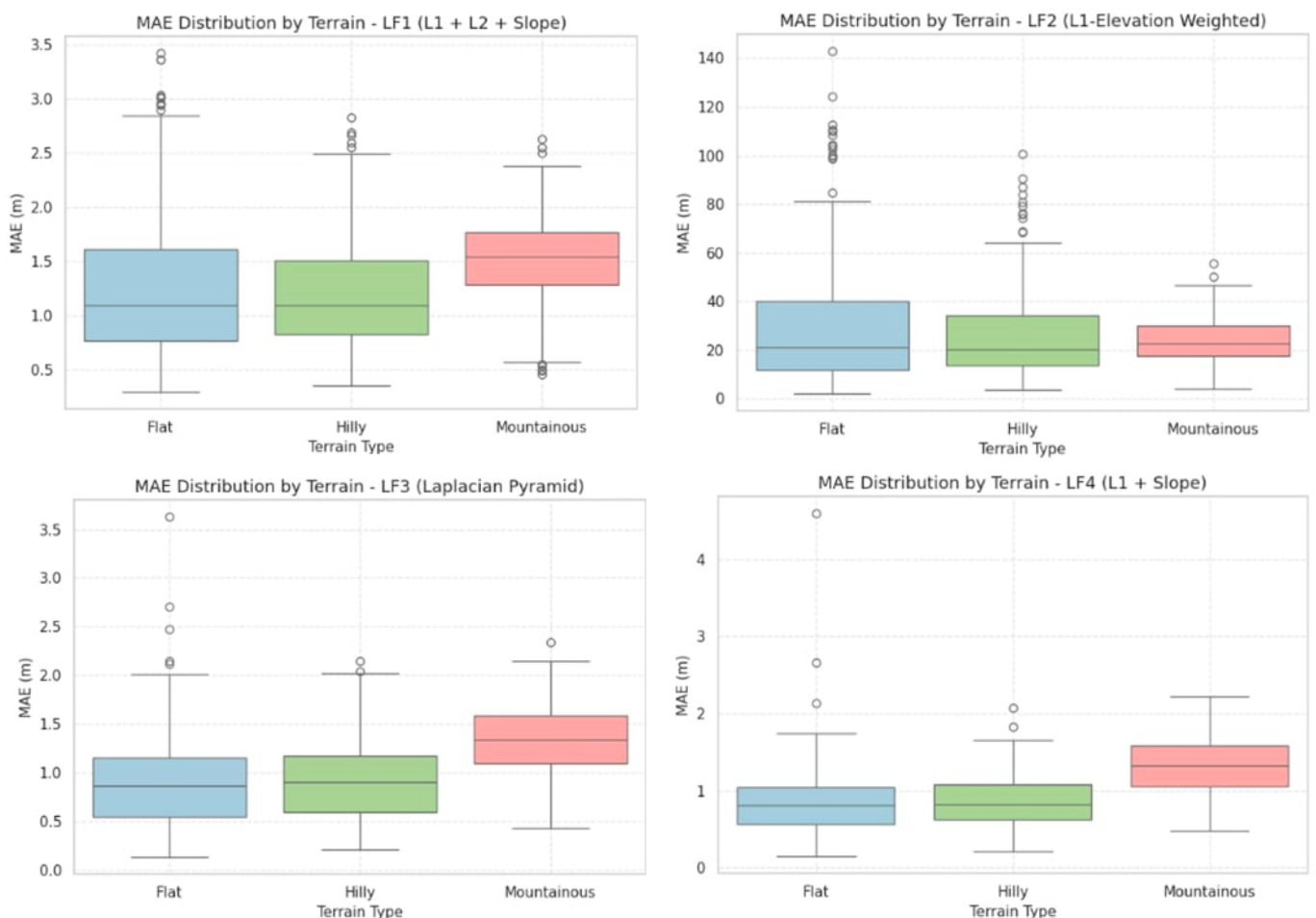


Figure 11. MAE distribution for Four loss functions.

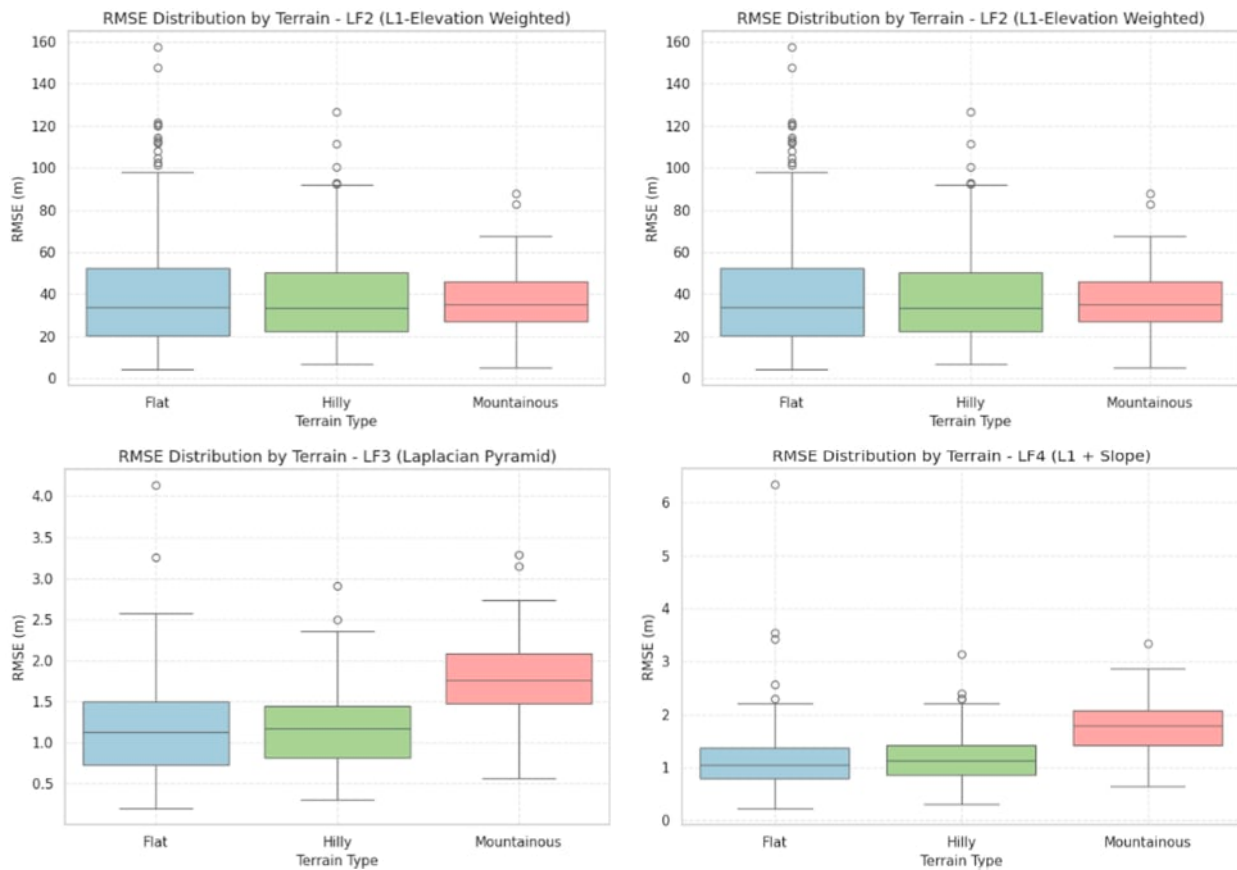


Figure 12. RMSE distribution for Four loss functions.

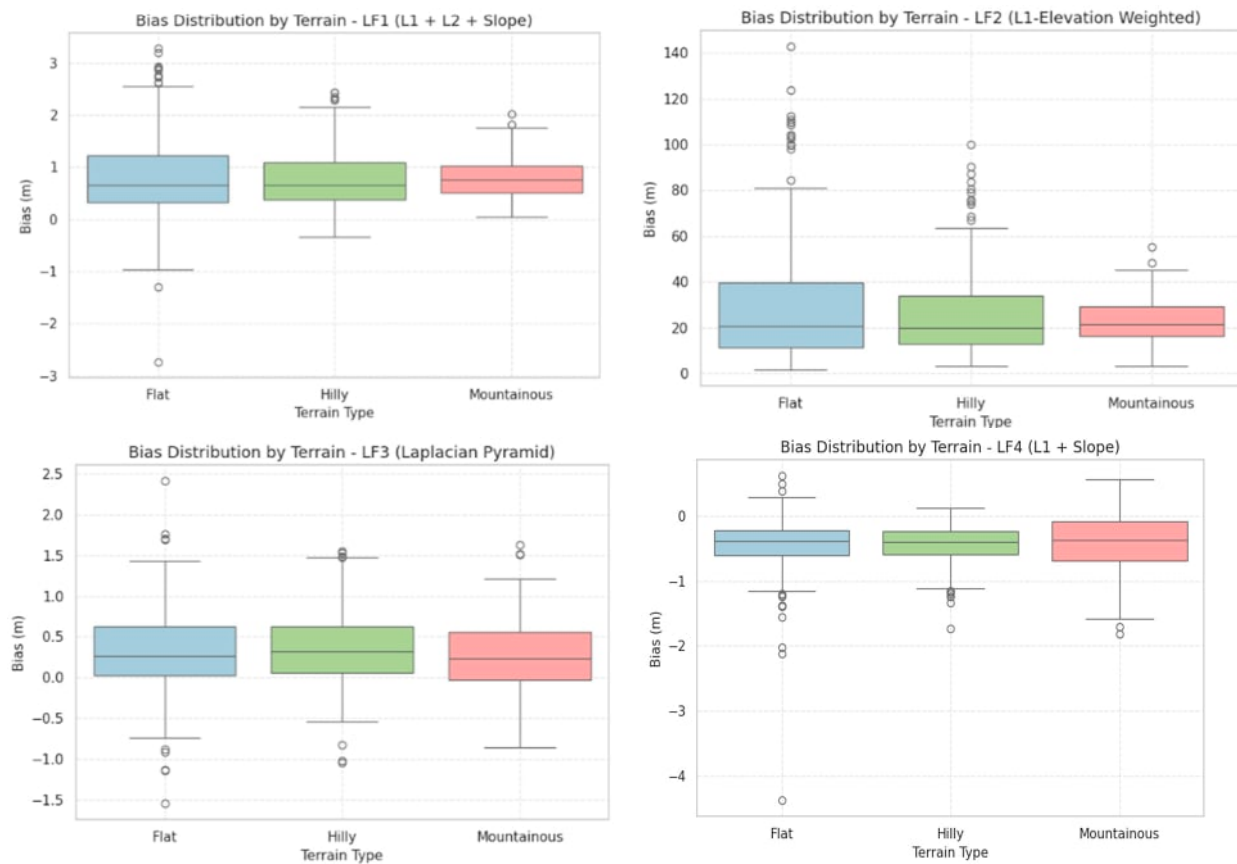


Figure 13. Bias distribution for Four loss functions.

Bias analysis revealed systematic patterns in elevation estimation. LF3 presented minimal bias across all terrain types (+0.26 m in mountainous areas), indicating strong balance and stability. LF1 consistently exhibited a positive bias (e.g., +0.82 m in flat regions), suggesting a slight overestimation trend. Conversely, LF4 showed a mild negative bias (−0.47 m in flat regions), indicating slight underestimation likely tied to its slope emphasis. LF2 displayed large positive biases in all regions (+29.76 m in flat terrain), confirming its instability.

3.4.2. Fine Tuning

To evaluate the performance of super-resolution models under different fine-tuning configurations for the LF4 loss function, the same accuracy metrics—MAE, RMSE, and Bias—were computed across flat, hilly, and mountainous regions. Results are summarized in Table 4 and visualized in Figures 14–16.

Table 4. Slope based zonal statistics assessment for different fine-tuning for LF4 loss function.

Model	Terrain	Mean MAE (m)	Mean RMSE (m)	Mean Bias (m)
$\alpha = 0.8, \gamma = 0.2$	Flat	0.85	1.15	−0.47
	Hilly	0.86	1.17	−0.43
	Mountainous	1.32	1.76	−0.41
$\alpha = 0.5, \gamma = 0.5$	Flat	0.83	1.14	−0.04
	Hilly	0.83	1.15	−0.03
	Mountainous	1.29	1.73	−0.04
$\alpha = 0.2, \gamma = 0.8$	Flat	1.02	1.29	+0.83
	Hilly	1.03	1.31	+0.80
	Mountainous	1.41	1.84	+0.72

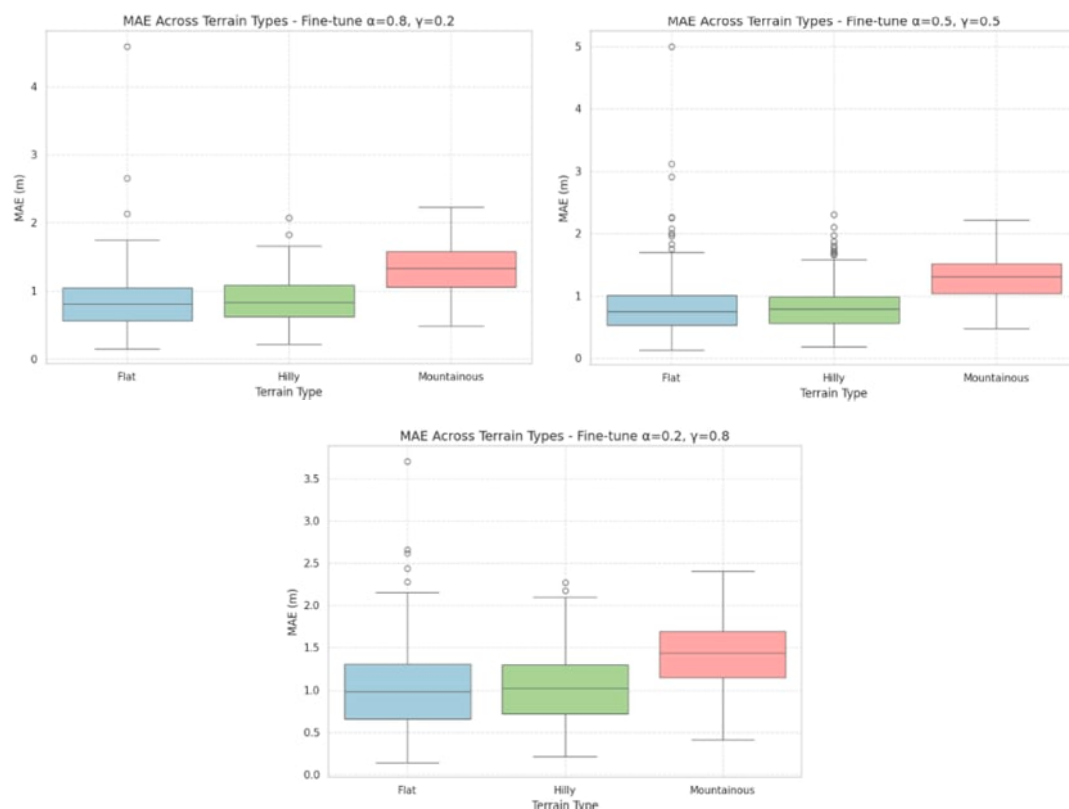


Figure 14. MAE distribution for Fine tuning.

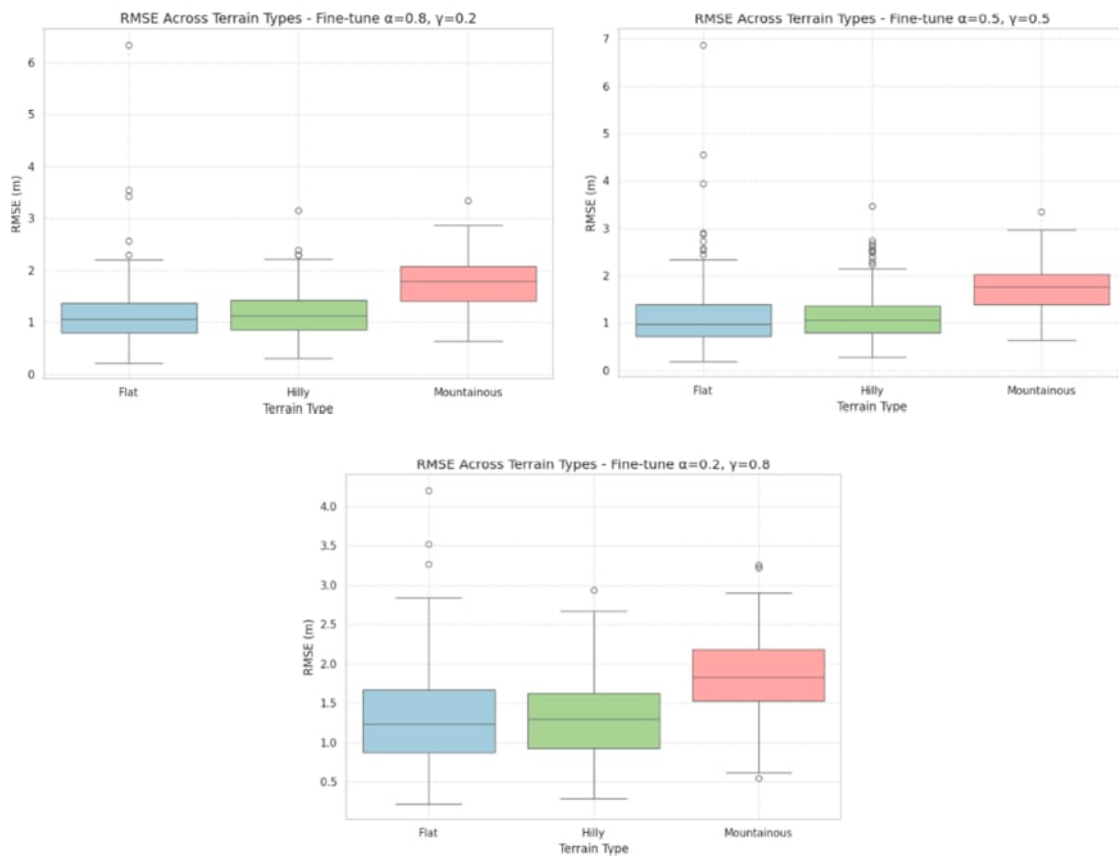


Figure 15. RMSE distribution for Fine tuning.

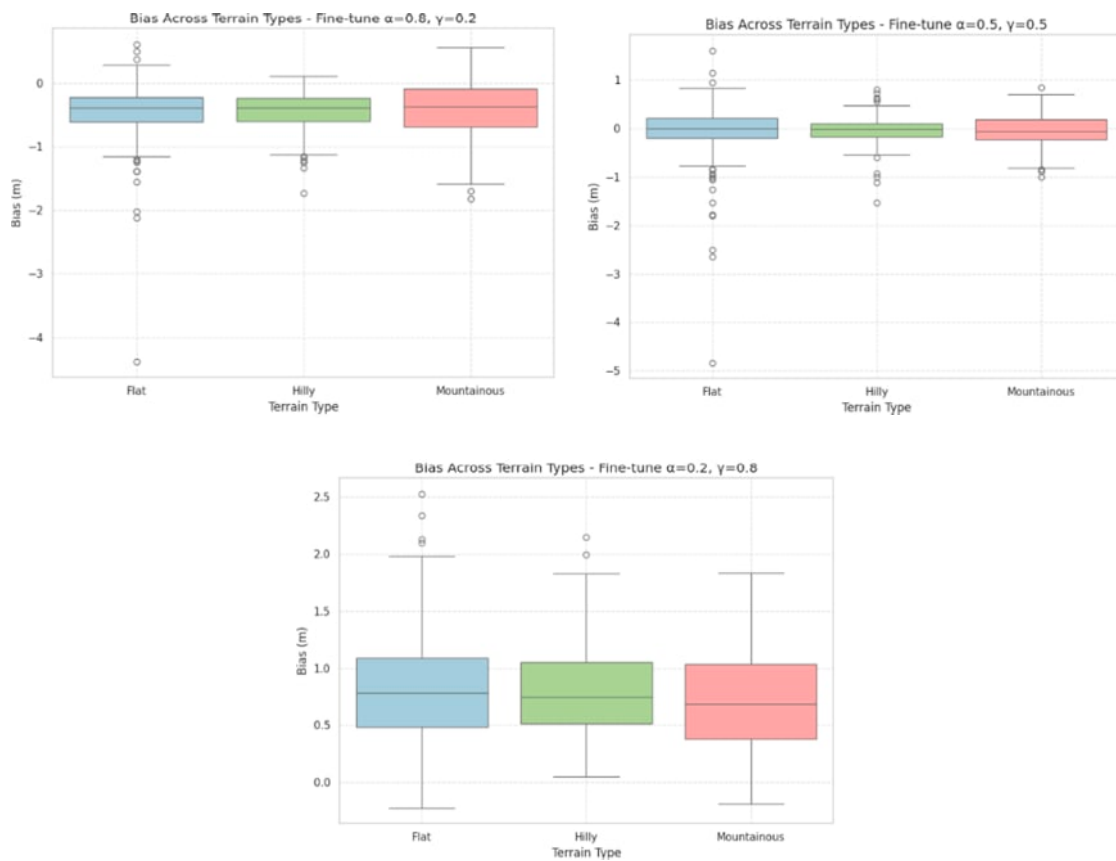


Figure 16. Bias distribution for Fine tuning.

MAE served as a key indicator of average elevation deviation. Across all terrain types, the lowest MAE values were consistently achieved in flat and hilly regions, where several fine-tuned models reached sub-meter accuracy. The best results were recorded in flat areas, with MAE as low as 0.83 m, indicating that fine-tuned LF4 configurations can closely approximate true elevation in homogeneous terrain. In contrast, mountainous regions exhibited higher MAE values, reaching up to 1.41 m, particularly when slope-based components dominated. Configurations balancing slope and elevation components achieved improved performance, with MAE around 1.29 m, suggesting better generalization in complex topography.

RMSE results mirrored the MAE trends. The lowest RMSE values appeared in flat and hilly areas (1.14–1.17 m), while mountainous regions showed higher errors (up to 1.84 m). RMSE proved most stable when slope and elevation weights were balanced, whereas disproportionate weighting toward either component led to greater variability and larger outlier errors. Balanced configurations not only reduced RMSE but also ensured consistent performance across all terrain classes.

Bias analysis revealed distinct systematic tendencies. Configurations emphasizing elevation accuracy exhibited negative bias (e.g., -0.47 m in flat regions), indicating mild underestimation of elevation. This pattern remained stable across terrains, suggesting that elevation-weighted (L1-dominant) losses compress elevation ranges. Conversely, slope-dominant configurations showed positive bias, particularly in flat areas ($+0.83$ m), reflecting overestimation due to excessive focus on gradient features. Balanced models maintained near-zero bias across all terrain types, achieving well-centered elevation predictions with minimal directional error.

3.5. Training Behavior and Convergence

3.5.1. Training Behavior of Different Loss Function

To assess the learning dynamics of the super-resolution model, the training loss was monitored over 100 epochs for each loss function (LF1–LF4). Figure 17 illustrates the training loss curves, which provide insight into convergence rate, stability, and overall learning behavior.

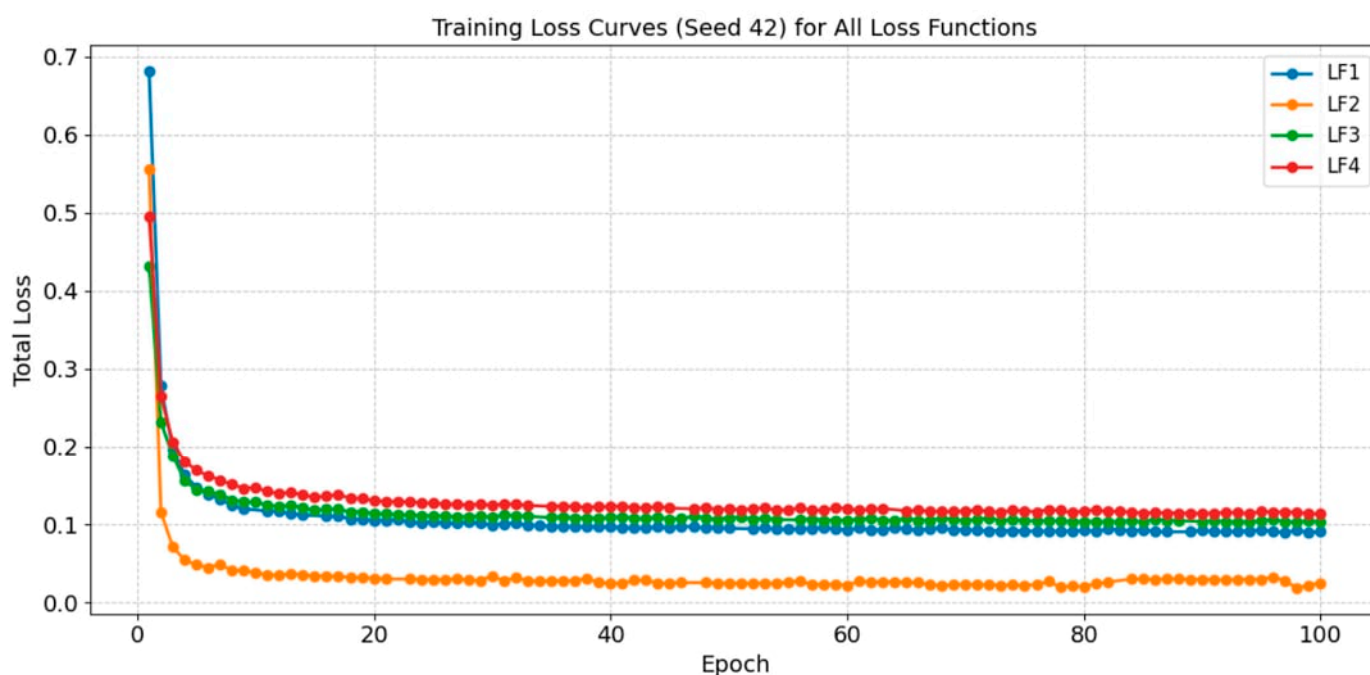


Figure 17. The training loss curves for the model trained with the four loss functions.

The results reveal distinct convergence characteristics among the four loss functions. LF1 demonstrated a rapid decline in loss during the initial epochs, stabilizing by approximately epoch 15 with a final loss value near 0.10. This pattern indicates efficient and consistent learning with no observable overfitting. LF2 exhibited a similar early trajectory, starting from a slightly lower initial loss and converging marginally faster to 0.06–0.07. Gentle oscillations in the later epochs suggest sustained fine-tuning, likely influenced by its elevation-weighted structure.

In contrast, LF3 showed a slower, more gradual convergence pattern, leveling off around epoch 30 at approximately 0.10. This behavior is consistent with the Laplacian-based formulation, which prioritizes structural preservation and evolves cautiously during optimization. LF4, which combines L1 and slope terms, displayed a steep initial loss reduction followed by a smooth, stable decline, reaching final values of 0.11–0.12. Despite converging to slightly higher loss levels, the curve remained notably stable, reflecting effective balance between accuracy and terrain-structure awareness.

3.5.2. Training Behavior Under Different Fine-Tuning Weights

To further optimize the model's capacity to preserve both elevation accuracy and terrain structure, the selected composite loss (LF4) was fine-tuned by varying the relative weights of its elevation (α) and slope (γ) components. The three configurations tested were ($\alpha = 0.8, \gamma = 0.2$), ($\alpha = 0.5, \gamma = 0.5$), and ($\alpha = 0.2, \gamma = 0.8$). The corresponding training loss curves are shown in Figure 18 and provide insight into how weighting affects convergence dynamics.

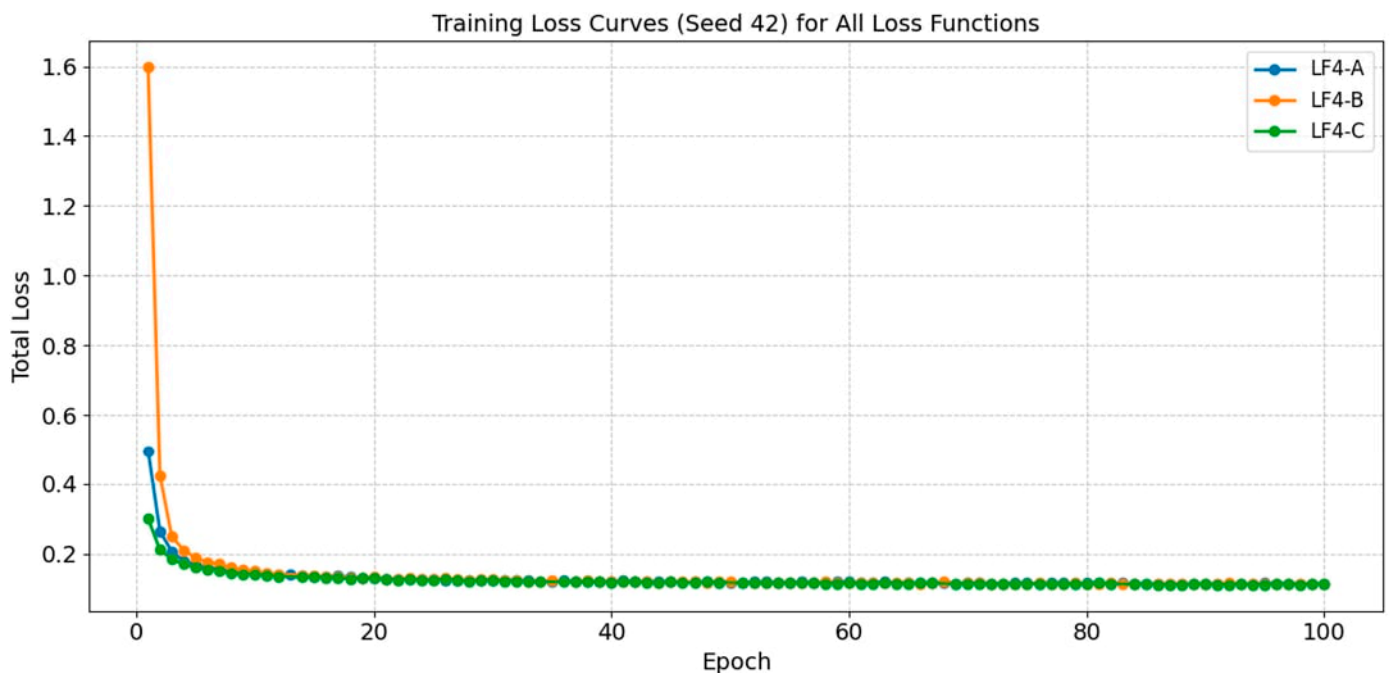


Figure 18. The training loss curves for the model trained with (A) ($\alpha = 0.8, \gamma = 0.2$), (B) ($\alpha = 0.5, \gamma = 0.5$), and (C) ($\alpha = 0.2, \gamma = 0.8$).

The elevation-dominant configuration ($\alpha = 0.8, \gamma = 0.2$) exhibited a rapid decline in training loss during the first few epochs, stabilizing around epoch 20 at a final value near 0.10. This behavior indicates strong convergence toward minimizing elevation errors. However, this configuration's weaker slope constraint may lead to reduced capability in preserving fine terrain structure in complex areas.

In contrast, the balanced configuration ($\alpha = 0.5, \gamma = 0.5$) showed slower but steadier convergence, reaching stability around epoch 30 and attaining the lowest final loss value of approximately 0.09. The smooth and consistent curve suggests an effective balance between minimizing elevation discrepancies and maintaining terrain morphology.

The slope-dominant configuration ($\alpha = 0.2, \gamma = 0.8$) converged more gradually, with the training loss flattening only after epoch 40 and reaching a slightly higher final value of about 0.12. Although this configuration enhances slope representation, especially in rugged regions, it converges more slowly and exhibits somewhat higher residual error in elevation prediction. Nevertheless, the absence of oscillations in all three cases confirms stable optimization throughout training.

4. Discussion

4.1. Visual Assessment of Terrain Reconstruction

The hill shade indicates that the proposed model enhances the visual interpretability of terrain by restoring coherent geomorphological patterns rather than merely increasing pixel density. Some smoothing is still evident in the super-resolved DTMs, particularly in steep or highly dissected areas, suggesting that certain fine-scale features remain challenging to reconstruct from coarse inputs. This behavior is consistent with the inherent information loss at 10 m resolution and reflects a trade-off between noise suppression and detail recovery.

4.2. Quantitative Evaluation of Loss Functions

The quantitative evaluation demonstrated that incorporating terrain-aware information into the loss function significantly enhances DTM super-resolution performance. Among the four tested formulations, the elevation-gradient loss (LF4) consistently outperformed the others across all key evaluation metrics, confirming the effectiveness of explicitly embedding slope information in the training process.

The superior performance of LF4 can be attributed to its balanced design, which simultaneously minimizes elevation discrepancies (through the L1 term) and preserves terrain morphology (via the slope-consistency term). This dual emphasis allows the model to maintain both vertical accuracy and horizontal structure, resulting in more realistic topographic reconstructions. In contrast, LF2's elevation-weighted approach introduced instability, suggesting that dynamically adjusting pixel importance based on elevation deviation may amplify noise or lead to overfitting, particularly in flat or uniform regions.

LF1 and LF3, while performing adequately, highlighted different strengths. LF1's composite structure improved general accuracy, but its reliance on multiple competing terms limited its ability to preserve sharp geomorphological transitions. LF3's Laplacian pyramid formulation better retained high-frequency details such as ridges and break lines, but its multi-scale weighting did not fully capture terrain continuity across slopes. The comparative results emphasize that simplicity and physical relevance in loss design—rather than mathematical complexity—yield better generalization for DTM enhancement tasks.

4.3. Loss Weight Tuning Analysis

The loss-weight tuning analysis highlights the critical role of maintaining equilibrium between elevation fidelity and slope consistency in terrain super-resolution. The superior performance of the balanced configuration ($\alpha = 0.5, \gamma = 0.5$) suggests that both elevation and gradient cues contribute complementary information: elevation terms ensure numerical stability, while slope terms preserve geomorphological continuity.

Overemphasizing the slope component ($\alpha = 0.2, \gamma = 0.8$) introduces excessive sensitivity to local gradients, which may amplify noise and degrade accuracy in smoother

landscapes. In contrast, relying too heavily on elevation alone ($\alpha = 0.8, \gamma = 0.2$) limits the model's ability to recover fine-scale structural detail, leading to smoother but less realistic reconstructions. The observed trends confirm that an intermediate weighting allows the model to generalize better across diverse terrain types by integrating both absolute and relational elevation information.

4.4. Terrain-Specific Performance

4.4.1. Loss Functions

The stratified terrain-based evaluation provides important insights into how topographic complexity affects the performance of different loss formulations. Both LF3 and LF4 performed robustly across all terrain classes, confirming that either integrating structural awareness, through multi-scale Laplacian components (LF3) or slope consistency (LF4), significantly improves the model's capacity to generalize beyond flat surfaces.

LF3's minimal bias and strong stability across varying relief suggest that its Laplacian pyramid component enhances sensitivity to multi-scale features while maintaining global elevation balance. This behavior is particularly beneficial for mountainous terrain, where abrupt elevation transitions can otherwise induce local distortions. LF4, although slightly biased toward underestimation, achieved the best overall numerical accuracy, indicating that slope-based constraints effectively preserve terrain geometry while limiting extreme errors.

In contrast, LF1, while consistent, exhibited a persistent positive bias, likely due to the dominance of pixel-wise losses (L1/L2) that do not explicitly encode structural gradients. Such models may produce smoother but slightly elevated terrain surfaces, potentially reducing hydrological realism. LF2's poor performance across all metrics highlights the limitations of elevation-weighted schemes that fail to integrate geometric regularization, resulting in large systematic offsets and reduced predictive reliability.

Most of the residuals reported in Table 3 are expected and typically occur in locations with abrupt high-frequency terrain changes, where fine-scale elevation discontinuities are difficult to fully reconstruct from 10 m inputs. These effects are well-known in DEM super-resolution and do not alter the overall performance trends.

4.4.2. Fine Tuning

The fine-tuning experiments for the LF4 loss function further highlight the critical role of balanced loss weighting in DTM super-resolution. The observed trends confirm that overemphasis on either slope or elevation terms compromises model generalization, particularly when transferring across terrains of varying relief.

In flat and hilly terrains, where gradients are gentle, slope-heavy models tend to over fit to minor elevation variations, resulting in artificial exaggeration of relief and positive bias. Conversely, elevation-dominant models underestimate heights, indicating a compression effect that smooths the terrain excessively. These systematic biases illustrate the inherent trade-off between preserving local geometry and maintaining global elevation accuracy.

The balanced configuration, however, mitigates these effects by jointly optimizing for structural integrity and elevation fidelity. The near-zero bias and sub-meter MAE observed across all terrain types indicate that the equal weighting scheme provides stable, terrain-independent performance. This balance allows the model to maintain sharp topographic features without amplifying noise or introducing elevation drift.

In mountainous terrain, where abrupt elevation changes and steep gradients dominate, even balanced models face increased RMSE due to the complexity of fine-scale relief. Nonetheless, their superior consistency across classes demonstrates robust generalization and a clear advantage for operational applications requiring uniform performance.

4.5. Training Behavior and Convergence

4.5.1. Training Behavior of Different Loss Function

The analysis of training dynamics highlights how loss design directly influences model convergence, learning stability, and overall optimization efficiency. The rapid and smooth convergence of LF1 and LF2 suggests that these formulations facilitate efficient gradient propagation, allowing the model to reach low loss values early in training. However, the slightly fluctuating pattern of LF2 indicates a potential trade-off: while its elevation-weighted loss accelerates learning, it may also introduce sensitivity to elevation range variability, potentially affecting stability across diverse terrains.

The gradual convergence of LF3 reflects the intrinsic behavior of Laplacian-based loss functions, which emphasize structural refinement and penalize abrupt spatial discrepancies. This slower learning pace is advantageous for preserving fine-scale terrain morphology, even though it delays reaching minimum loss values.

Meanwhile, LF4 achieves a desirable compromise between learning speed and stability. Its training curve demonstrates controlled, consistent convergence without oscillations or divergence—an indicator of robust optimization. The slightly higher final loss value does not necessarily imply inferior performance; rather, it reflects a more conservative adjustment process, which prioritizes slope and gradient consistency alongside elevation accuracy.

4.5.2. Training Behavior Under Different Fine-Tuning Weights

The convergence trends across the three fine-tuning configurations highlight how loss weighting directly governs learning behavior and optimization stability in DTM super-resolution. The rapid convergence observed in the elevation-heavy configuration ($\alpha = 0.8$, $\gamma = 0.2$) indicates efficient gradient flow when elevation accuracy dominates the training objective. However, this setup tends to prioritize vertical precision over geomorphological realism, which may reduce structural fidelity in complex terrains.

Conversely, the slope-heavy configuration ($\alpha = 0.2$, $\gamma = 0.8$) emphasizes surface gradients and morphological features but does so at the expense of absolute elevation accuracy and convergence speed. The slower training observed for this setup suggests that learning detailed slope patterns requires more epochs to stabilize, particularly in less variable regions such as plains.

The balanced configuration ($\alpha = 0.5$, $\gamma = 0.5$) emerges as the most stable and efficient compromise. It maintains smooth convergence, achieves the lowest final training loss, and effectively integrates both elevation and slope learning objectives. This equilibrium allows the model to preserve fine-scale morphological detail without introducing significant elevation bias or instability.

The results also point toward several avenues in which the current framework could be extended. The present study focuses on elevation-aware loss functions that encode slope information to preserve geomorphic structure, but the architecture could be expanded into a multi-task formulation. Predicting elevation and slope jointly, rather than deriving slope as a secondary product, may strengthen the physical consistency of the output and reduce structural drift in steep or heterogeneous terrain. A second direction concerns model confidence. Terrain super-resolution models often behave differently across landforms, and an explicit estimate of uncertainty—whether through dropout sampling, ensemble strategies, or other probabilistic methods—would help quantify how the model reacts in areas dominated by sharp breaks or smooth plains. These ideas do not alter the core findings but highlight where the approach could evolve as higher-resolution training material becomes available.

5. Conclusions

This study introduced a deep learning-based framework for the super-resolution of Digital Terrain Models (DTMs) using a modified Residual Channel Attention Network (RCAN) architecture. The model successfully upscaled 10 m input DTMs to 1 m resolution and was trained and evaluated on a LiDAR-derived dataset covering diverse Italian landscapes, including flat, hilly, and mountainous regions. Central to the methodology was the design and optimization of elevation-aware loss functions, combining absolute elevation accuracy (L1) with slope-preserving terms to improve terrain realism and precision.

Experimental results demonstrated that a balanced loss configuration—equally weighting elevation and slope components—provided the most robust and generalizable performance. This setting achieved the lowest RMSE and MAE values across all terrain classes, while maintaining a bias close to zero, indicating well-centered predictions without systematic over- or underestimation. Although the achieved accuracy remains slightly above the nominal precision of the LiDAR reference data, the improvements are substantial, especially in complex terrains. Mountainous regions presented higher variability and error magnitudes, yet the model effectively preserved key geomorphological features, confirming its ability to retain structural integrity even in high-relief conditions.

Analysis of training behavior further revealed that Laplacian- and slope-aware losses enhanced structural learning at the cost of slower convergence, while simpler elevation-based losses converged faster but with reduced morphological fidelity. The study also highlights that models biased toward either elevation or slope accuracy alone tend to introduce directional prediction errors, whereas balanced configurations maintain both accuracy and stability.

Despite its strong performance, the framework's generalization to other regions, resolutions, or input sources such as SRTM or ALOS remains to be validated. Additionally, the computational cost and the need for multiple training seeds to ensure stability may constrain large-scale or real-time applications. Nonetheless, the model's lightweight inference and demonstrated reliability position it as a promising tool for geomorphology, hydrology, and landscape analysis, particularly in areas lacking LiDAR coverage.

Beyond the limitations identified here, several technical directions could strengthen the physical consistency and reliability of future terrain super-resolution frameworks. A multi-task formulation—where the network jointly predicts elevation and slope instead of deriving slope as a secondary product—may help stabilize geometric continuity in areas with abrupt gradients. Likewise, incorporating uncertainty estimation, such as ensemble predictions or dropout-based sampling, would offer insight into where the model is confident and where its outputs require caution. These additions do not alter the main conclusions of this study but point toward meaningful extensions for improving model robustness across diverse geomorphological settings.

Future research will focus on extending the model's transferability to diverse terrains and sensor inputs, incorporating uncertainty quantification, and refining its performance toward near-LiDAR-level precision. With continued optimization, this framework has strong potential to become an operational approach for enhancing elevation data quality across a wide range of environmental applications.

6. Limitations and Practical Implications

The super-resolution model developed in this study demonstrates strong potential for enhancing Digital Terrain Models (DTMs) from coarse to fine resolution; however, several limitations constrain its broader applicability. The model was trained exclusively on LiDAR-derived DTMs from selected regions of Italy, covering terrain types such as flat plains, hilly areas, and mountainous zones. Consequently, its generalization to other

geographic contexts or data sources, including global DTMs such as SRTM or ALOS World 3D, remains unconfirmed; however, we are currently exploring the application of our model to these datasets, aiming to assess its transferability and performance in a wider geographic context. The model may also exhibit reduced reliability in terrain types absent from the training dataset, such as volcanic fields, dune systems, densely built-up environments, or periglacial areas, where elevation patterns differ significantly from those encountered during training. Extending the training dataset to include morphologies that are more diverse such as volcanic fields, dune systems, densely built-up environment, or periglacial environments would likely improve its adaptability to these terrains.

Furthermore, while the framework effectively supports a $10\times$ upscaling factor (10 m–1 m), extending its use to coarser datasets (e.g., 30 m–3 m) would likely require architectural modifications, retraining, or fine-tuning to preserve both elevation accuracy and structural integrity. A critical limitation lies in the model's dependence on high-quality ground truth data. LiDAR, although precise, is costly, geographically limited, and computationally demanding to process, restricting the scalability of this approach to larger or less surveyed regions.

Training the RCAN-based model also proved computationally intensive, with convergence achieved only after 100 epochs across multiple loss configurations and random seeds. The combination of large tile sizes and the $10\times$ upsampling ratio further increased GPU memory and processing requirements, posing challenges for large-scale or repeated applications. Additionally, comparisons with other state-of-the-art super-resolution models, such as SRGAN, EDSR, or recent Transformer-based architectures, were not performed due to their limited upsampling capabilities relative to RCAN. Including such comparisons represents a valuable future direction for a more comprehensive evaluation of the model's performance.

Despite these constraints, the proposed framework shows strong potential for operational use in fields such as geomorphology, hydrology, and terrain analysis, particularly where LiDAR data are unavailable. Applications may include drainage network extraction, slope stability assessment, and landform characterization, where enhanced elevation precision can substantially improve analytical outcomes. For practical deployment, it is essential to validate the model across diverse environmental conditions and data sources, ensuring robustness and adaptability. Future work should also incorporate uncertainty quantification mechanisms to identify potential quality issues and guide users in interpreting model outputs confidently. By addressing these aspects, the framework could evolve into a reliable and transferable tool for global-scale terrain enhancement and geospatial analysis.

Author Contributions: M.H.: Conceptualization, Methodology, Software, Field study, Data curation, Writing—Original draft preparation, Software, Validation. and Field study; E.M., L.V. and G.B.: Visualization, Investigation, Writing—Reviewing and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study are publicly available. The LiDAR-DTM reference data were provided by the Italian Ministry for the Environment and Energy Security (MASE) and are available upon request from the data provider.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Guth, P.L.; Geoffroy, T.M. LiDAR point cloud and ICESat-2 evaluation of 1 second global digital elevation models: Copernicus wins. *Trans. GIS* **2021**, *25*, 2245–2261. [[CrossRef](#)]

2. Passalacqua, P.; Do Trung, T.; Fofoula-Georgiou, E.; Sapiro, G.; Dietrich, W.E. A geometric framework for channel network extraction from lidar: Nonlinear diffusion and geodesic paths. *J. Geophys. Res. Earth Surf.* **2010**, *115*. [[CrossRef](#)]
3. Tarolli, P.; Mudd, S.M. *Remote Sensing of Geomorphology*; Elsevier: Amsterdam, The Netherlands, 2020; Volume 23.
4. Ackroyd, C.; Skiles, S.M.; Rittger, K.; Meyer, J. Trends in snow cover duration across river basins in High Mountain Asia from daily gap-filled MODIS fractional snow covered area. *Front. Earth Sci.* **2021**, *9*, 713145. [[CrossRef](#)]
5. Bertin, S.; Jaud, M.; Delacourt, C. Assessing DEM quality and minimizing registration error in repeated geomorphic surveys with multi-temporal ground truths of invariant features: Application to a long-term dataset of beach topography and nearshore bathymetry. *Earth Surf. Process. Landf.* **2022**, *47*, 2950–2971. [[CrossRef](#)]
6. James, M.R.; Quinon, J.N. Ultra-rapid topographic surveying for complex environments: The hand-held mobile laser scanner (HMLS). *Earth Surf. Process. Landf.* **2014**, *39*, 138–142. [[CrossRef](#)]
7. Lehner, S.; Pleskachevsky, A.; Velotto, D.; Jacobsen, S. Meteo-marine parameters and their variability: Observed by high-resolution satellite radar images. *Oceanography* **2013**, *26*, 80–91. [[CrossRef](#)]
8. Mukherjee, S.; Joshi, P.K.; Mukherjee, S.; Ghosh, A.; Garg, R.D.; Mukhopadhyay, A. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 205–217. [[CrossRef](#)]
9. Atkinson, P.M. Downscaling in remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *22*, 106–114. [[CrossRef](#)]
10. Xu, Z.; Wang, X.; Chen, Z.; Xiong, D.; Ding, M.; Hou, W. Nonlocal similarity based DEM super resolution. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 48–54. [[CrossRef](#)]
11. Zhang, Y.; Yu, W. Comparison of DEM super-resolution methods based on interpolation and neural networks. *Sensors* **2022**, *22*, 745. [[CrossRef](#)]
12. Zhou, T.; Geng, Y.; Chen, J.; Pan, J.; Haase, D.; Lausch, A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. *Sci. Total Environ.* **2020**, *729*, 138244. [[CrossRef](#)] [[PubMed](#)]
13. Tsai, R.Y.; Huang, T.S. Multiframe image restoration and registration. *Multiframe Image Restor. Regist.* **1984**, *1*, 317–339.
14. Yang, W.; Zhang, X.; Tian, Y.; Wang, W.; Xue, J.H.; Liao, Q. Deep learning for single image super-resolution: A brief review. *IEEE Trans. Multimed.* **2019**, *21*, 3106–3121. [[CrossRef](#)]
15. Zhou, A.; Chen, Y.; Wilson, J.P.; Su, H.; Xiong, Z.; Cheng, Q. An enhanced double-filter deep residual neural network for generating super resolution DEMs. *Remote Sens.* **2021**, *13*, 3089. [[CrossRef](#)]
16. Lin, X.; Zhang, Q.; Wang, H.; Yao, C.; Chen, C.; Cheng, L.; Li, Z. A DEM super-resolution reconstruction network combining internal and external learning. *Remote Sens.* **2022**, *14*, 2181. [[CrossRef](#)]
17. Yan, L.; Fan, B.; Xiang, S.; Pan, C. CMT: Cross mean teacher unsupervised domain adaptation for VHR image semantic segmentation. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 6002205. [[CrossRef](#)]
18. Kim, K.I.; Kwon, Y. Single-image super-resolution using sparse regression and natural image prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1127–1133.
19. Sun, J.; Xu, Z.; Shum, H.Y. Gradient profile prior and its applications in image super-resolution and enhancement. *IEEE Trans. Image Process.* **2010**, *20*, 1529–1542. [[CrossRef](#)]
20. Chang, H.; Yeung, D.Y.; Xiong, Y. Super-resolution through neighbor embedding. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004), Washington, DC, USA, 27 June–2 July 2004; IEEE: New York, NY, USA, 2004; Volume 1, p. I.
21. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [[CrossRef](#)]
22. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [[CrossRef](#)]
23. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HA, USA, 21–26 July 2017; pp. 4681–4690.
24. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HA, USA, 21–26 July 2017; pp. 136–144.
25. Winiwarter, L.; Mandlbürger, G.; Schmohl, S.; Pfeifer, N. Classification of ALS point clouds using end-to-end deep learning. *PFG–J. Photogramm. Remote Sens. Geoinf. Sci.* **2019**, *87*, 75–90. [[CrossRef](#)]
26. Zhang, Y.; Yu, W.; Zhu, D. Terrain feature-aware deep learning network for digital elevation model superresolution. *ISPRS J. Photogramm. Remote Sens.* **2022**, *189*, 143–162. [[CrossRef](#)]
27. Sun, J.; Xu, F.; Cervone, G.; Gervais, M.; Wauthier, C.; Salvador, M. Automatic atmospheric correction for shortwave hyperspectral remote sensing data using a time-dependent deep neural network. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 117–131. [[CrossRef](#)]

28. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
29. Jiang, Y.; Xiong, L.; Huang, X.; Li, S.; Shen, W. Super-resolution for terrain modeling using deep learning in high mountain Asia. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103296. [[CrossRef](#)]
30. Qiu, Z.; Yue, L.; Liu, X. Void filling of digital elevation models with a terrain texture learning model based on generative adversarial networks. *Remote Sens.* **2019**, *11*, 2829. [[CrossRef](#)]
31. Feng, R.; Grana, D.; Mukerji, T.; Mosegaard, K. Application of Bayesian generative adversarial networks to geological facies modeling. *Math. Geosci.* **2022**, *54*, 831–855. [[CrossRef](#)]
32. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [[CrossRef](#)]
33. Liu, K.; Ding, H.; Tang, G.; Song, C.; Liu, Y.; Jiang, L.; Zhao, B.; Gao, Y.; Ma, R. Large-scale mapping of gully-affected areas: An approach integrating Google Earth images and terrain skeleton information. *Geomorphology* **2018**, *314*, 13–26. [[CrossRef](#)]
34. Zhou, S.; Feng, Y.; Li, S.; Zheng, D.; Fang, F.; Liu, Y.; Wan, B. DSM-assisted unsupervised domain adaptive network for semantic segmentation of remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5608216. [[CrossRef](#)]
35. Rasera, L.G.; Gravey, M.; Lane, S.N.; Mariethoz, G. Downscaling images with trends using multiple-point statistics simulation: An application to digital elevation models. *Math. Geosci.* **2020**, *52*, 145–187. [[CrossRef](#)]
36. Bertero, M.; Boccacci, P.; De Mol, C. *Introduction to Inverse Problems in Imaging*; CRC Press: Boca Raton, FL, USA, 2021.
37. Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: Oxford, UK, 1997.
38. Kyriakidis, P.C.; Shortridge, A.M.; Goodchild, M.F. Geostatistics for conflation and accuracy assessment of digital elevation models. *Int. J. Geogr. Inf. Sci.* **1999**, *13*, 677–707. [[CrossRef](#)]
39. Mariethoz, G.; Renard, P.; Straubhaar, J. The direct sampling method to perform multiple-point geostatistical simulations. *Water Resour. Res.* **2010**, *46*. [[CrossRef](#)]
40. Remy, N.; Boucher, A.; Wu, J. *Applied Geostatistics with SGeMS: A User's Guide*; Cambridge University Press: Cambridge, UK, 2009.
41. Strebelle, S. Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* **2002**, *34*, 1–21. [[CrossRef](#)]
42. Straubhaar, J.; Renard, P.; Mariethoz, G. Conditioning multiple-point statistics simulations to block data. *Spat. Stat.* **2016**, *16*, 53–71. [[CrossRef](#)]
43. Zhang, T.; Switzer, P.; Journel, A. Filter-based classification of training image patterns for spatial simulation. *Math. Geol.* **2006**, *38*, 63–80. [[CrossRef](#)]
44. Ke, L.; Song, C.; Wang, J.; Sheng, Y.; Ding, X.; Yong, B.; Ma, R.; Liu, K.; Zhan, P.; Luo, S. Constraining the contribution of glacier mass balance to the Tibetan lake growth in the early 21st century. *Remote Sens. Environ.* **2022**, *268*, 112779. [[CrossRef](#)]
45. Atwood, A.; West, A.J. Evaluation of high-resolution DEMs from satellite imagery for geomorphic applications: A case study using the SETSM algorithm. *Earth Surf. Process. Landf.* **2022**, *47*, 706–722. [[CrossRef](#)]
46. Musa, Z.N.; Popescu, I.; Mynett, A. A review of applications of satellite SAR, optical, altimetry and DEM data for surface water modelling, mapping and parameter estimation. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 3755–3769. [[CrossRef](#)]
47. Deng, Y.; Wilson, J.P.; Bauer, B.O. DEM resolution dependencies of terrain attributes across a landscape. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 187–213. [[CrossRef](#)]
48. Shary, P.A.; Sharaya, L.S.; Mitusov, A.V. The problem of scale-specific and scale-free approaches in geomorphometry. *Geogr. Fis. E Din. Quat.* **2005**, *28*, 81–101.

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.