remote sensing

MDPI

Article

A Feature-Fusion Convolutional Neural Network for Terrain Generalization

Yanchen Wan [1], Xinyue Lyu [1], Han Jiang [2], Yue Wang [1], Wenping Jiang [1,*], Jiawei Fan [1] and Haowen Jiang [1]

[1] School of Resource and Environmental Sciences, Wuhan University, 129 LuoYu Road, Wuhan 430079, China; yanchenwan@whu.edu.cn (Y.W.)
[2] Dornsife College of Letters, Arts and Sciences, University of Southern California, 3551 Trousdale Pkwy, Los Angeles, CA 90089, USA; hjiang53@usc.edu
* Correspondence: andyjiang@whu.edu.cn

Highlights

What are the main findings?

- A feature-fusion CNN (TG-CNN) is built for 3× DEM terrain downsampling, with max-/average-pooling downsampling variants and slope plus circularly encoded aspect embedded in both the network inputs and a composite loss.
- A multi-faceted comparison with classical methods—bilinear and nearest-neighbor resampling, TIN-based simplification, and IDW—characterizes TG-CNN in terms of both elevation accuracy (MAE/RMSE) and ridge–valley structure. The interpolation methods attain the highest elevation accuracy, the fused TG-CNN+IDW scheme benefits chiefly from its IDW component, and TG-CNN suppresses secondary ridge–valley branches more strongly than the other methods.

What are the implications of the main findings?

- TG-CNN achieves a genuine terrain generalization effect, retaining the dominant ridge–valley skeleton while eliminating secondary microtopography—evidence that the network performs scale-driven abstraction rather than mere elevation fitting.
- In terms of numerical elevation error, however, TG-CNN does not surpass classical interpolation, revealing a practical limitation of deep learning in terrain generalization and indicating that further research should be directed toward objectives beyond low-ratio elevation accuracy.

Abstract

Terrain generalization—the scale-driven abstraction of digital elevation models (DEMs) that retains dominant landforms while filtering trivial microtopography—has long been dominated by rule-based algorithms, whereas deep learning has mainly targeted the inverse task of DEM super-resolution. This study proposes TG-CNN, a feature-fusion convolutional neural network for DEM down-scaling generalization. The network compresses the high-resolution DEM through a max- or average-pooling front end, combines normalized elevation with slope and sine–cosine circularly encoded aspect as complementary input channels, and is trained with a composite elevation–slope–aspect loss. Training, validation, and testing are conducted on geographically independent mountainous areas under a leave-one-area-out scheme, followed by adjustable IDW fusion post-processing. On the test region, classical bilinear interpolation retains the best elevation accuracy, while the standalone TG-CNN does not reach a comparable level; fused accuracy improves monotonically as the IDW share increases, indicating that the fusion benefit originates from the interpolation component. On the ridge–valley structure, however, TG-CNN behaves as an

Check for updates

Academic Editor: Gino Dardanelli

Received: 12 August 2026
Revised: 2 October 2026
Accepted: 20 October 2026
Published: 23 October 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.

aggressive yet reliable filter: it achieves the highest precision and the smallest positional offset among all compared methods but retains only part of the reference skeleton, leaving its F1-score below the interpolation baseline. TG-CNN thus realizes a genuine, controllable terrain-abstraction effect, yet deep learning does not yet surpass mature interpolation in pixel-level elevation accuracy—a negative result that clarifies current limitations and points toward residual learning and information-preserving architectures.

Keywords: terrain generalization; digital elevation model; deep learning; convolutional neural network; feature fusion

1. Introduction

Terrain constitutes the fundamental land surface form of the Earth, and its accurate visualization and information transmission have long posed a core challenge in geographic information science. For multi-scale geospatial analysis, terrain generalization [1]—the deliberate, rule-governed abstraction and simplification of continuous, complex terrain according to map scale, landform type, or research purpose—serves as the foundation for building multi-scale geospatial databases and conducting diverse geoscientific analyses. This process focuses on information abstraction rather than simple data compression; it not only influences the visual legibility and information richness of maps [2] but also governs the precision of hydrological simulation, landscape ecology analysis, military terrain evaluation, and other geospatial applications. Accordingly, developing reliable terrain generalization techniques has remained a central research topic in cartography and geographic information science, driving continuous technical innovation within this discipline.

Conventional terrain generalization approaches fall into two primary groups: image-processing-based algorithms and terrain-structure-based algorithms. Weibel [3] pointed out that cartographic generalization is restricted by multiple constraints, including terrain morphology, geographic feature types, map scale, graphical specifications, and data quality. Appropriate processing pipelines must be selected following analysis of terrain structures and application workflows. As such, traditional terrain generalization heavily relies on expert experience and fixed pre-defined simplification rules.

From a task perspective, it is necessary to clarify the essential differences among grid downsampling, low-resolution DEM reconstruction and cartographic terrain generalization. Straightforward downsampling only downsizes raster grids via mathematical aggregation such as pooling or averaging, which merely compresses data volume without the hierarchical screening of geomorphic features. Low-resolution DEM reconstruction mainly focuses on pixel-value fitting to generate coarse grids, lacking cartographic constraints for preserving complete ridge-and-valley topological frameworks. By contrast, terrain generalization is a scale-driven cartographic abstraction activity that retains primary landform skeletons and filters trivial microtopographic fluctuations under geographic rules, which fundamentally differentiates it from pure resampling and pixel-level reconstruction.

In recent years, deep learning, particularly convolutional neural networks (CNNs), has offered new perspectives to address this long-standing challenge in terrain generalization. Advances in CNN and Generative Adversarial Network (GAN)-based DEM super-resolution have fully demonstrated that deep neural networks can capture complex cross-scale geomorphic spatial correlations between coarse and fine terrain grids. Nevertheless, DEM super-resolution and cartographic terrain generalization are fundamentally distinct tasks and cannot be regarded as simple inverse operations. DEM super-resolution follows a low-to-high-resolution mapping direction, aiming to infer and restore all subtle microtopographic fluctuations to enrich fine-scale terrain details. By contrast, terrain gener-

alization implements purposeful geomorphic information abstraction from high-resolution inputs: it retains primary ridge–valley skeletons and eliminates trivial secondary terrain undulations in compliance with map scale, representation goals, and cartographic constraints.

A standard super-resolution model optimized only by pixel-wise elevation loss lacks geomorphic topological supervision. As typical black-box architectures, these models cannot autonomously judge which terrain features should be retained or removed during simplification and frequently generate visually plausible yet spatially shifted or artificial ridge and valley structures. Low pixel-level elevation error cannot guarantee the accurate spatial positioning of core geomorphic elements such as drainage divides, channel networks, and critical terrain feature points. Accordingly, the latent representations learned by vanilla super-resolution models cannot be directly adapted to scale-dependent terrain abstraction. Despite abundant publications on DEM super-resolution, few studies have developed network architectures and loss functions specifically tailored to meet the geomorphic–topological consistency requirements of terrain generalization.

Against this background, this study moves beyond the limitations of conventional rule-based terrain generalization and unconstrained DEM super-resolution networks and proposes TG-CNN, a feature-fusion convolutional neural network for knowledge-guided terrain down-scaling. Rather than directly reusing existing super-resolution architectures without targeted redesign, we tailor the network structure and multi-objective loss function to fit the core demands of cartographic terrain simplification. For experimental validation, we adopt two types of mountainous study areas separately from China and the United States and conduct a fixed 3× downsampling task to comprehensively verify the robustness and cross-regional generalization ability of the proposed model. The main contributions of this study are summarized as follows:

(1) A geomorphology-aware CNN framework specifically developed for 3× terrain down-scaling is established. Two alternative spatial compression modules—max pooling and average pooling—are separately configured and compared, providing an experimental basis for selecting the downsampling strategy best suited to terrain generalization.

(2) Geomorphic prior knowledge is integrated into the model through multi-channel inputs and a composite loss function. Slope and circularly encoded sine–cosine aspect features are fed into the network as auxiliary inputs, and corresponding gradient loss terms are added to the objective function to drive scale-consistent terrain abstraction. A leave-one-area-out cross-validation scheme using geographically independent training and validation regions is adopted to evaluate cross-region generalization.

(3) A multi-dimensional comprehensive evaluation workflow coupled with tunable IDW post-processing is constructed to overcome the overreliance on single elevation-error metrics prevalent in existing DEM deep learning research. The adjustable IDW fusion module enables continuous control of terrain generalization intensity without retraining the network. The evaluation system unifies pixel-wise MAE (mean absolute error)/RMSE (root mean square error) quantitative measurement, hillshade visual interpretation and ridge–valley skeleton extraction analysis, enabling joint assessment of numerical elevation precision and the cartographic rationality of simplified landform structures under different generalization strengths.

2. Related Work

2.1. Traditional Terrain Generalization Methods

Traditional terrain generalization methods are predominantly image-processing-based and include resampling, low-pass filtering, and wavelet analysis [1]. Resampling simplifies data by reducing pixel count. Its core is performing numerical aggregation such as mean value and bilinear interpolation on original DEM pixels using local neighborhood windows

such as 3 × 3 and 5 × 5. It can quickly reduce the data volume. Low-pass filtering is a traditional numerical smoothing method for DEM generalization. It uses fixed-size windows to convolve DEM grid data, suppressing high-frequency terrain details and small undulations while preserving low-frequency main terrain contours, thus achieving terrain smoothing and data simplification. Wavelet analysis, derived from the fields of signal processing, is a mathematical method that can decompose data into different frequency components at a specific resolution to match multi-scale expression needs and can be used to extract edge information [4]. As a nonlinear filtering technology, wavelet analysis is adaptive, which preserves large-scale wavelet coefficients in steep terrain and discards sub-threshold coefficients in flat areas. It can preserve local important terrain structures.

Weibel [3] systematically reviewed terrain generalization methods, dividing them into three categories: global filtering methods, selective filtering methods, and structure-based heuristic methods. Resampling and low-pass filtering both belong to global filtering methods. Selective filtering methods assess the importance of DEM points by using specified rules [5], retain high-importance points, and remove less important points to generate lower-resolution DEMs [6]. They are primarily used for TIN (Triangulated Irregular Network) generalization and include the very important points (VIP) method [7], the 2D and 3D Douglas–Peucker algorithms [8], and point addition and subtraction methods [9,10].

Since DEM data contain various terrain features, such as ridges, peaks, saddles, valleys, and plains, DEM generalization is widely regarded as information abstraction rather than simple data reduction [1]. Therefore, in special scenarios that require high-quality DEMs that preserve geomorphological features while maintaining a smooth representation, image-processing methods are insufficient, and terrain-structure-based methods are preferred [3]. Terrain-structure-based DEM generalization generally follows a three-stage workflow. Terrain features such as ridges and valleys are first detected and extracted from the original DEM [11–14]. Following feature extraction, the hierarchical importance of each terrain structure is quantitatively assessed to distinguish primary and secondary landform elements [15–17]. After low-significance terrain features are screened out, coarse-resolution generalized DEMs are reconstructed by eliminating secondary structural lines. Ai et al. [1] simplified DEMs by filling secondary valleys, but linear filling led to insufficient local smoothness. Chen et al. [18] improved the VIP method using the D8 algorithm and tree-model-based feature extraction, achieving good performance and efficiency. Chen et al. [19] proposed a graph-Laplacian-based TIN method that preserved more structural details.

2.2. Application of Machine Learning in Terrain Processing

Machine learning, especially deep learning, has brought a new paradigm to terrain analysis with its strong data-driven feature learning ability. Many scholars have applied deep learning to various terrain analyses, confirming the feasibility and superiority of deep learning in learning terrain features. In terms of terrain feature extraction and semantic understanding, Zhao et al. [20] used Generative Adversarial Networks (GANs) to fill DEM data gaps and used ridge lines, valley lines, and DEM data as network inputs. The results showed that this method overcame the limitations in reconstruction accuracy of complex terrain and consistency of cavity boundaries and was suitable for DEM data of different resolutions and sources. Bittner et al. [21] designed an end-to-end model architecture for generating DTM data from DSM, and the results maintained the steep terrain features of mountains. Yang et al. [22] automatically performed pixel-level classification of basic geomorphological types using only 30 m resolution DEM data, revealing that sufficient geomorphological semantic information can be extracted using the original DEM. Ho et al. [23] realized high-precision, gap-free DTM construction by fusing multi-source data and a global–local transfer learning framework and generated terrain and hydrological

parameters at six scales. Syzdykbayev et al. [24] automatically extracted key terrain features from shaded relief maps of digital terrain models, and performed better than or comparable to existing mainstream methods both on synthetic datasets and in real scenarios.

In terms of DEM super-resolution reconstruction, studies have confirmed that models such as CNNs and GANs can effectively learn the complex nonlinear mapping from low-resolution DEMs to high-resolution DEMs and restore delicate terrain textures [25]. Chen et al. [26] first applied a CNN to the field of DEM reconstruction and achieved good results. Since then, many scholars have continuously refined model architectures and input features to improve reconstruction accuracy, such as introducing terrain features [27,28], changing convolutional layers [29], introducing remote sensing images [30], etc., all of which have confirmed the advantages of deep learning in the field of DEM reconstruction. In particular, the MTF-SR model proposed by Zhou et al. [28] achieved strong performance in terms of elevation accuracy and terrain feature preservation, which indicates that deep networks can implicitly learn rich terrain feature information and demonstrates the potential of data-driven methods in terrain expression.

In the field of deep learning-based DEM processing, existing studies have predominantly concentrated on super-resolution reconstruction, which aims to recover fine-scale terrain details from low-resolution inputs. This body of research has thoroughly demonstrated that convolutional neural networks are capable of capturing complex cross-scale geomorphic spatial correlations. However, DEM super-resolution and cartographic terrain generalization are fundamentally distinct tasks with opposing mapping directions and optimization objectives. Vanilla super-resolution CNNs represent typical black-box architectures, optimized solely by pixel-wise elevation loss without embedded geomorphic or cartographic constraints. Even when pixel-level elevation errors appear low, these models tend to generate artificial ridge–valley structures and spatially misaligned drainage divides, as their latent feature representations lack explicit supervision of terrain hierarchy and hydrological connectivity. Such learned representations cannot be directly applied to scale-dependent terrain abstraction. To date, few studies have constructed dedicated deep learning pipelines to address the specific demands of terrain generalization, and existing works have not systematically resolved the mismatch between super-resolution feature learning and cartographic simplification rules.

2.3. Summary

In summary, cartographic terrain generalization research has long been dominated by traditional rule-based algorithms, while deep learning techniques for DEM processing are predominantly focused on super-resolution up-scaling tasks. As a result, dedicated deep learning frameworks tailored specifically for DEM down-scaling generalization remain relatively underdeveloped. Both established research branches present well-documented, unaddressed limitations that motivate this work.

This paper fills the two-tiered research gap outlined above via the proposed framework, TG-CNN, where every core design is purpose-built to tackle the respective drawbacks of existing methods. To overcome the deficiencies of traditional rule/threshold-based terrain generalization, we introduce slope paired with sine–cosine circularly encoded aspect as auxiliary input channels and formulate a composite loss function integrating elevation loss and terrain gradient loss. By embedding geomorphic prior knowledge throughout model training, the framework effectively safeguards the topological continuity of ridge-and-valley skeletons during terrain simplification.

To mitigate feature misalignment and artificial terrain artifacts inherent to unconstrained black-box DEM super-resolution models, we develop two parallel downsampling variants based on max pooling and average pooling for comparative analysis.

The customized max-pooling module dynamically suppresses insignificant microtopographic signals while emphasizing prominent core landform structures, reorienting the feature learning logic from super-resolution reconstruction to scale-aware cartographic terrain abstraction.

In addition, we construct a multi-dimensional quantitative evaluation framework consisting of MAE, RMSE and topological metrics derived from extracted ridge–valley lines. This system jointly assesses pixel-wise elevation precision and the geomorphological rationality of simplified DEMs, remedying the overreliance on single elevation-error metrics prevalent in prior deep learning terrain research.

3. Methodology

3.1. A Multi-Terrain Feature-Based Deep CNN for Super-Resolution (SR) DEMs (MTF-SR)

MTF-SR, proposed by Zhou et al. [28], is a residual network tailored for DEM super-resolution reconstruction. Stacked residual blocks are adopted within the model to suppress gradient disappearance and explosion during training, which enables stable extraction of subtle topographic features from grid DEM data and achieves reliable retention of fine terrain textures.

The robust feature extraction capacity of its residual structure inspired our network design. On this basis, we independently construct a dedicated convolutional neural network, TG-CNN, to fulfill the demand of multi-scale terrain generalization.

3.2. Terrain Generalization Convolutional Neural Network (TG-CNN)

The residual architecture of the existing MTF-SR model delivers stable extraction of subtle terrain features and serves as a structural reference for this work, yet MTF-SR is originally built for low-to-high DEM super-resolution upsampling. By redesigning its overall pipeline to fit high-to-low cartographic terrain generalization, we propose the fully customized feature-fusion network TG-CNN, whose complete architecture is illustrated in Figure 1.

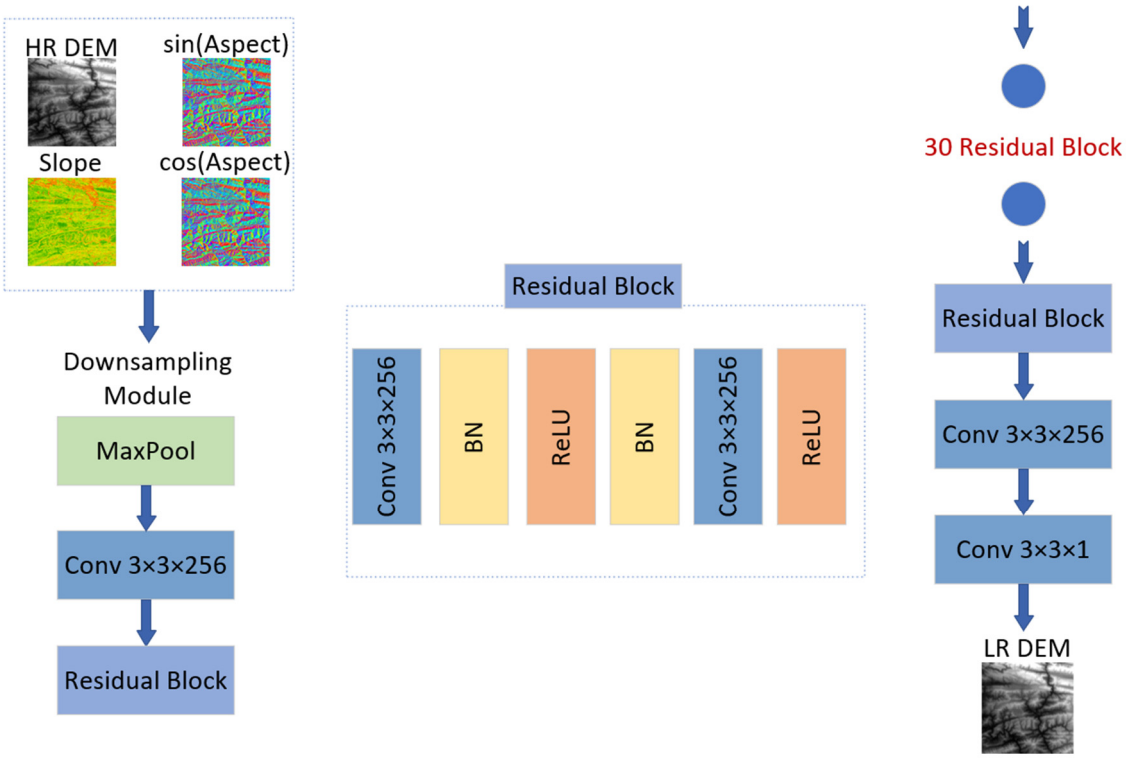


Figure 1. The network architecture of TG-CNN (HR: high resolution; LR: low resolution).

We further customize the residual building block: batch normalization layers are inserted on both sides of the ReLU activation within each residual unit. This design continuously regularizes intermediate feature distributions and mitigates violent gradient fluctuations, substantially improving network convergence stability.

The network accepts a four-channel input tensor. The four input channels are preprocessed terrain features: normalized elevation, standardized slope, sine-encoded aspect and cosine-encoded aspect. The input tensor first passes through the front end 3×3 max-pooling module to realize 3-fold spatial downsampling. Afterwards, an input convolution layer maps the 4-channel input into 256-dimensional feature space for geomorphic feature excavation. The high-dimensional feature maps are then fed into 32 stacked improved residual blocks for multi-level terrain-skeleton feature representation. A feature-fusion convolution layer compresses the feature dimension to 64, followed by ReLU activation. Finally, a single-output convolution layer outputs the normalized low-resolution DEM prediction.

To satisfy the core generalization demand of retaining major geomorphic skeletons while removing trivial micro-undulations, we construct a front end max-pooling downsampling unit for spatial compression, with its operating logic being displayed in Figure 2. Max pooling is a standard untrainable CNN operator, and our task-oriented customizations lie in two targeted designs: for the $3\times$ downsampling task in this study, the kernel size and stride are strictly set to 3 to guarantee uniform spatial scaling across all feature channels. The inherent maximum-value screening characteristic naturally highlights ridge peaks and valley bottoms during compression, consistent with the skeleton-retention principle of cartographic abstraction.

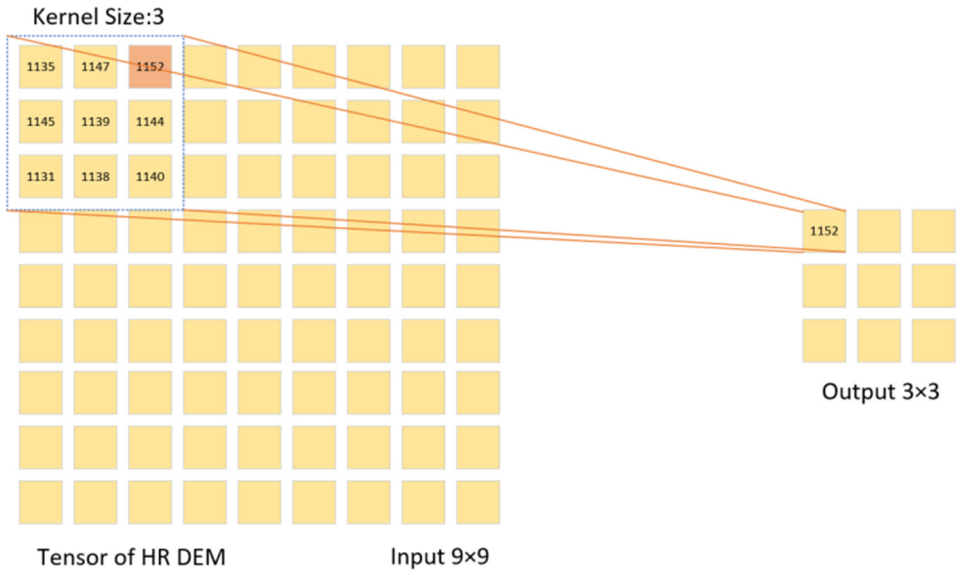


Figure 2. The principle of the maximum pooling layer for DEM downsampling.

For comparative analysis, an alternative configuration replacing the max-pooling downsampling module with average pooling is also constructed. All other architectural components—including the residual blocks, feature fusion, and output layers—remain identical. The two pooling variants are evaluated and compared in Section 4.3.

Slope and circularly encoded aspect features are computed via central finite differences from input DEM tensors. Let c denote the ground sampling distance of the input DEM. The partial derivatives along two orthogonal directions are calculated as shown in Equation (1):

$$dx = \frac{z_{i,j+1} - z_{i,j-1}}{2c}, \; dy = \frac{z_{i+1,j} - z_{i-1,j}}{2c} \tag{1}$$

where dx and dy represent elevation gradients in the east–west and north–south directions, respectively. The slope θ (in radians) is defined as shown in Equation (2):

$$\theta = \arctan(\sqrt{dx^2 + dy^2}) \tag{2}$$

To avoid numerical instability of the atan2 function under flat terrain, sine- and cosine-encoded aspect are derived directly from gradient components.

A small epsilon $\varepsilon = 10^{-10}$ is added to prevent division by zero:

$$m = \sqrt{dx^2 + dy^2} + \varepsilon \tag{3}$$

$$\sin(\phi) = \frac{dx}{m},\ \cos(\phi) = \frac{dy}{m} \tag{4}$$

where $\sin(\phi)$ and $\cos(\phi)$ are the circularly encoded aspect channels with value range $[-1, 1]$.

3.3. Training Strategy of TG-CNN

Raw elevation, slope and aspect cannot be directly fed into the network due to inconsistent value magnitudes. Tile-wise pre-processing and normalization are applied to construct the four-channel input features. Tile-wise Z-score normalization is applied to elevation following Equation (5):

$$E_{norm} = \frac{E - \mu}{\sigma} \tag{5}$$

where E denotes raw elevation, and μ and σ represent the mean and standard deviation of each DEM tile. Slope is linearly scaled to [0, 1] by dividing its arctangent result by $\pi/2$. Instead of using a single normalized angular channel suffering from the well-known 0–360° circular discontinuity, aspect is encoded as two circular components, $\sin(\varphi)$ and $\cos(\varphi)$ (Equations (3) and (4)), generating two independent input channels ranging in $[-1, 1]$.

3.3.1. Multi-Objective Loss Function

We adopt a multi-objective composite loss to balance elevation fitting and terrain structural consistency. The elevation loss (Equation (6)) is defined as the mean absolute error of normalized DEM outputs in Z-score space, while the slope loss (Equation (7)) calculates absolute deviation in terrain gradient at the target resolution using the de-normalized predictions in physical units. For aspect, a linear angular difference would penalize the 1–359° pair by 358° instead of the true 2° separation. We therefore quantify circular aspect error through joint sine–cosine component deviation computed from the de-normalized predictions (Equation (8)):

$$L_{elev} = \frac{1}{N}\sum_{i=1}^{N}\left|E_{norm,pred,i} - E_{norm,true,i}\right| \tag{6}$$

$$L_{slope} = \frac{1}{N}\sum_{i=1}^{N}\left|S_{pred,i} - S_{true,i}\right| \tag{7}$$

$$L_{aspect} = \frac{1}{2N}\sum_{i=1}^{N}\left(\left|A_{sin,pred,i} - A_{sin,true,i}\right| + \left|A_{cos,pred,i} - A_{cos,true,i}\right|\right) \tag{8}$$

The total weighted loss function (Equation (9)):

$$L_{total} = L_{elev} + \alpha L_{slope} + \beta L_{aspect} \tag{9}$$

The weighting coefficients $\alpha = 0.6$ and $\beta = 0.0$ are determined to balance the numerical ranges of the three loss components to comparable magnitudes, ensuring that no single term dominates gradient propagation during training. This choice is validated through a parameter-sensitivity ablation experiment (Section 4.3).

3.3.2. Training Configuration

In each training epoch, 200 independent patches are randomly sampled with a batch size of 2. Random horizontal and vertical flipping are applied for data augmentation. For horizontal flip, the circularly encoded aspect components are updated as $\sin(\varphi)_{\text{new}} = -\sin(\varphi)_{\text{old}}$ and $\cos(\varphi)_{\text{new}} = \cos(\varphi)_{\text{old}}$. For vertical flip, the aspect components are transformed as $\sin(\varphi)_{\text{new}} = \sin(\varphi)_{\text{old}}$ and $\cos(\varphi)_{\text{new}} = -\cos(\varphi)_{\text{old}}$. Such component-wise updates maintain physically consistent directional semantics for the two-channel circularly encoded aspect after geometric transformation.

A leave-one-area-out cross-validation scheme is employed to ensure spatial independence between training and validation data. One complete geographic region (e.g., Valdez) serves as the training set, while an entirely separate geographic region (e.g., Chongqing) is used for validation. This design ensures that the validation metric reflects genuine cross-region generalization rather than within-tile interpolation between spatially adjacent patches. No DEM tile is used simultaneously for both training and validation. After each training epoch, 80 patches are randomly sampled from the validation region to compute the validation loss.

The network is optimized using the Adam [31] optimizer with an initial learning rate of 5×10^{-5} and a weight decay of 1×10^{-5}. A cosine-annealing warm-restart scheduler ($T_0 = 20$, T_mult = 2, and η_min = 1×10^{-6}) dynamically adjusts the learning rate to help the model escape local minima. Per-value gradient clipping (clip_value = 0.1) is adopted to prevent gradient explosion within the deep residual blocks. Early stopping with a patience of 15 epochs monitors the validation loss computed on the held-out geographic region; training terminates if the validation loss shows no continuous decline over 15 consecutive epochs. The model weights and tile-wise normalization parameters corresponding to the lowest validation loss are saved for subsequent inference.

3.3.3. IDW Post-Processing Refinement

After model inference, inverse distance weighting (IDW) serves as a post-processing refinement to recover subtle terrain details that the pixel-wise L_1 loss tends to smooth. A KD-tree spatial index is constructed over the nodes of the 30 m high-resolution (HR) grid, and for each low-resolution (LR) pixel center, the K = 12 nearest HR nodes are retrieved. Because one 90 m cell covers a 3×3 block of 30 m nodes, K = 12 spans this block together with its first ring of neighbors, keeping the interpolation local enough to preserve detail while remaining stable when an LR center coincides with an HR node. Distances are measured on the HR grid in pixel units (1 unit = 30 m), and a half-pixel floor $d_k \leftarrow \max(d_k, 0.5)$ is applied to avoid division by zero. The refined elevation is computed as:

$$d_k = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \tag{10}$$

$$z_{\text{IDW}} = \frac{\sum_{k=1}^{M} \frac{z_k}{d_k^p}}{\sum_{k=1}^{M} \frac{1}{d_k^p}} \tag{11}$$

$$Z_{\text{enhanced}} = \alpha_{\text{idw}} \cdot Z_{\text{pred}} + (1 - \alpha_{\text{idw}}) \cdot Z_{\text{IDW}} \tag{12}$$

where d_k represents the Euclidean distance between the interpolation point $(x_i y_i)$ and the k-th high-resolution reference pixel. z_{IDW} represents the refined elevation from inverse distance weighting interpolation, M is the total number of high-resolution reference pixels in interpolation, z_k represents the elevation of the k-th high-resolution reference pixel, and p is the power exponent controlling the influence of distance on weight ($p = 2$ in this study). A larger p gives more weight to nearby pixels and makes the interpolation more localized. $z_{enhanced}(i,j)$ represents the finally optimized elevation (at pixel (i, j)). α is the fusion weight coefficient (usually 0–1) to balance model prediction and interpolation, and $z_{pred}(i,j)$ represents the predicted elevation from the deep learning model (at pixel (i, j)).

The optimal fusion coefficient α_idw is determined through a parameter-sensitivity experiment (Section 4.4), where α_idw is varied across {0.0, 0.3, 0.5, 0.7, 1.0} and the resulting MAE and RMSE are compared against the reference low-resolution DEM.

3.4. Study Areas

3.4.1. Training Areas

The experiments employ two geographically independent mountainous study areas (Figure 3) with consistent 3× down-scaling configurations, namely, a mountainous region in Chongqing, China, and the Valdez multi-resolution DEM dataset from southern Alaska, USA [32]. The Chongqing area adopts a uniform dimension of 1800 × 1800 pixels, while each Valdez tile measures 1500 × 1500 pixels. Only native, real-measured 30 m high-resolution (HR) and 90 m low-resolution (LR) DEM pairs are adopted for the 3-fold terrain generalization task in this work, without artificially synthesizing low-resolution DEM from single high-resolution data; other redundant resolution layers of the Valdez dataset are not utilized. The Chongqing study area covers the geographic extent ranging from 108°30′E to 109°E and from 31°N to 31°30′N, located at the junction of the eastern Sichuan hills and the residual veins of the Daba Mountains. This area features typical high mountains and deep gullies with drastic undulation, with elevation distributed between 102 m and 2233 m, and its 30 m and 90 m DEMs are independent real products derived from SRTM surveys. The Valdez multi-resolution DEM (Version 1.0, released on 14–19 June 2019) was compiled by retired US NPS researcher Tom Patterson, adopting the Albers equal-area projection based on the NAD83 geodetic datum, with standard parallels of 55°N and 65°N, an origin latitude of 50°N and a central meridian of 146.5°W. Centered at 61.1428366°N, 146.371743°W, this area belongs to the middle section of the Alaska Range, characterized by steep folded fault mountains with severe terrain fluctuation. Its 30 m and 90 m DEMs are two separate real datasets sourced from USGS 3DEP IFSAR measurements, and the complete multi-resolution series of Valdez was generated via bicubic interpolation during collation without additional artificial generalization filtering. All elevation materials for the Valdez dataset are open-access public domain resources released under the Creative Commons Attribution 4.0 International license. Known data defects including abrupt coastal elevation jumps and north–south stitching seams near the Columbia Glacier, as well as NoData pixels, coastline and marginal areas in both study areas, are uniformly masked during pre-processing to exclude abnormal elevation values from model training and evaluation.

To avoid spatial patch overlap and data leakage, each complete DEM tile from the two regions is geographically partitioned into mutually exclusive training, validation and test subdomains with no adjacent or intersecting coverage. Then, 384 × 384 pixel sub-regions are randomly cropped only from the training partition during sampling, and the spatially independent validation set serves as the sole metric for the early-stopping strategy rather than training loss, strictly following standardized deep learning experimental specifications. Both study areas are dominated by mountain landforms interspersed with scattered

flat valley plains, providing abundant and diversified terrain samples covering ridges, valleys and transitional zones of varying relief. The dual-region design effectively prevents the model from overfitting to a single geomorphic type, and the unified pairing of real 30 m and 90 m DEMs across two distinct mountainous zones enables reliable cross-terrain verification of model performance. The fundamental metadata of the two experimental areas are listed in Table 1.

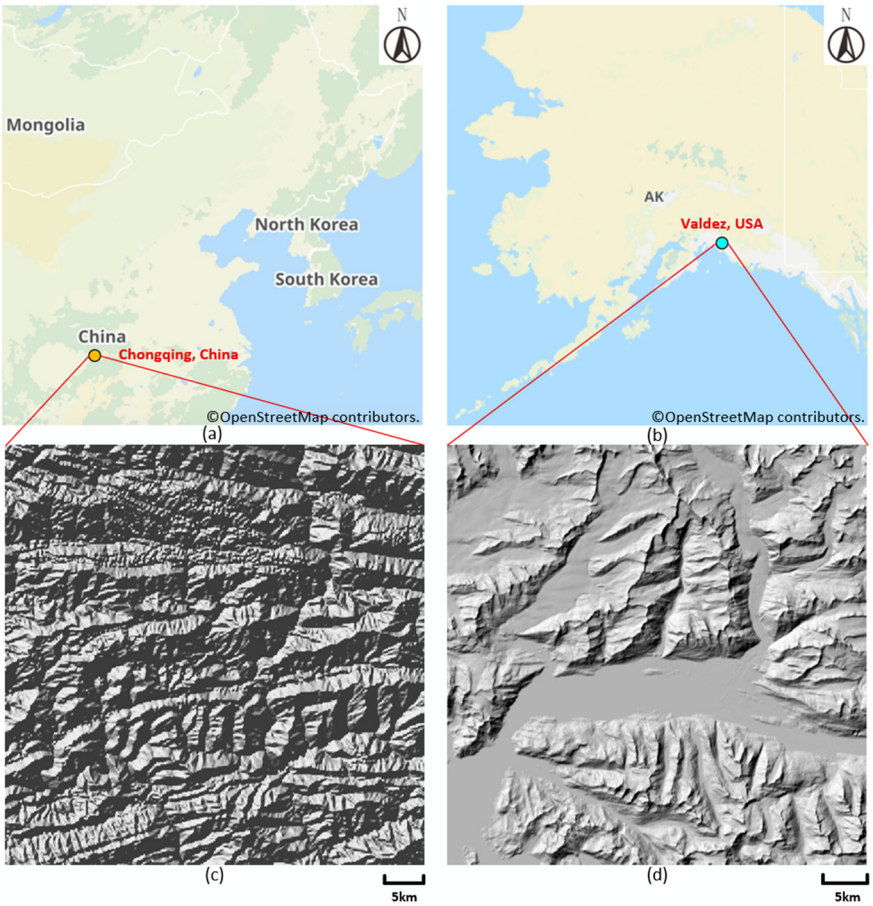


Figure 3. Training areas: (**a**) Location of Chongqing. (**b**) Location of Valdez. (**c**) Hillshade map of Chongqing (30 m). (**d**) Hillshade map of Valdez (30 m).

Table 1. Training area details.

Region	Geographical Range	Elevation Range	Resolutions
Chongqing	108°30′E–109°E, 31°N–31°30′N	102 m–2233 m	30 m, 90 m
Valdez	146°47′W–145°56′W, 60°56′N–61°20′N	−2.3 m–2599 m	30 m, 90 m

3.4.2. Test Area

The independent test region adopts the Gore Range (Figure 4) multi-resolution DEM dataset (Version 1.0), compiled by retired US NPS researcher Tom Patterson and released on 4 May 2020. The dataset adopts the Albers equal-area projection based on the NAD83 geodetic datum, with standard parallels of 29.5°N and 45.5°N, an origin latitude of 23°N, and a central meridian of 106°W. Each DEM tile has a fixed size of 1500 × 1500 pixels, with the geographic center at 39.7377625°N, 106.3282392°W. Consistent with the two training areas, only native and independently surveyed 30 m high-resolution and 90 m low-resolution DEMs are selected for 3× down-scaling, and other resolution layers contained in this dataset are not involved in our experiments. The 30 m DEM of the Gore Range mainly

originates from 1 m LiDAR measurements, and peripheral areas are supplemented by 3.3 m LiDAR data, while its 90 m DEM comes from separate USGS 3DEP observation data. In this study, none of the low-resolution samples are generated by downsampling a single high-resolution DEM.

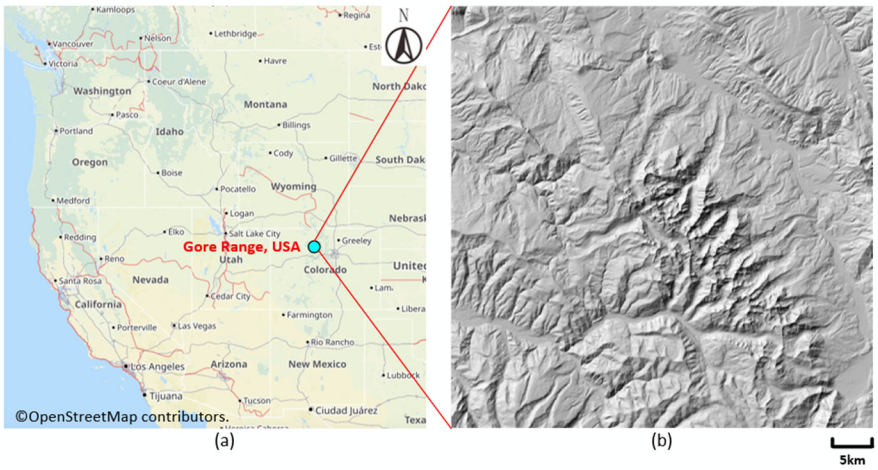


Figure 4. Test area: (**a**) Test area location. (**b**) Hillshade map of Gore Range (30 m).

The complete multi-resolution DEM library of the Gore Range was assembled via Natural Scene Designer Pro, QGIS and Geographic Imager for mutual conversion among NSD, IMG, ASCII Grid and TIFF formats. Bicubic interpolation was merely adopted for auxiliary image splicing and format unification during overall dataset production, rather than being used to generate the 30 m and 90 m DEM pairs applied in our training and test pipelines. No extra terrain smoothing or generalization filtering was added in the dataset collation process. Minor vertical scan artifacts appear on partial talus slopes of the 1 m ultra-fine DEM, which will not interfere with the 30 m and 90 m data adopted in this work. All NoData and edge pixels are masked in the pre-processing stage to eliminate abnormal elevation samples before quantitative evaluation. The whole dataset is released under the Creative Commons Attribution 4.0 International license, and its underlying elevation data from USGS 3DEP belong to public domain resources.

Different from the Chongqing and Valdez training areas, the Gore Range is a geographically separated alpine zone completely unseen during model training. It features unique topographic features, which can objectively measure the cross-landform generalization performance of the proposed network, TG-CNN, and avoid biased evaluation results induced by terrain overfitting.

4. Results

4.1. Experimental Setup and Baseline Methods

This section adopts the well-tuned hyper-parameter set of the proposed network, TG-CNN, where the optimized configuration is specified as follows: the Valdez study area for model training, a max-pooling downsampling unit, loss weight coefficients $\alpha = 0.6$ and $\beta = 0.0$, and an IDW fusion coefficient of 0.3. Detailed parameter-sensitivity analyses comparing pre-optimization and post-optimization hyper-parameter settings for individual hyper-parameters are presented in subsequent sections. Quantitative elevation-based evaluation (Table 2) is carried out with the high-resolution reference DEM (HR), the ground-truth low-resolution DEM (LR), a TIN-based DEM simplification method [33] guided by geomorphological structural lines [34], bilinear resampling, and TG-CNN (before and after hyper-parameter optimization). The comparison focuses on elevation-related metrics to examine the capability of the proposed model to preserve fundamental terrain characteristics.

Table 2. Comparison of elevation.

Method	Maximum Elevation (m)	Minimum Elevation (m)	Average Elevation (m)
HR	4127.10	2075.42	2953.31
LR	4127.44	2075.55	2952.84
TG-CNN	4090.13	2255.37	2962.89
TG-CNN + IDW (α_{idw} = 0.3)	4108.80	2131.32	2955.87
Bilinear	4119.98	2075.51	2953.30
TIN-based	4114.42	2075.86	2953.74

Figure 5 shows the elevation frequency distribution histograms of DEMs derived from all tested methods.

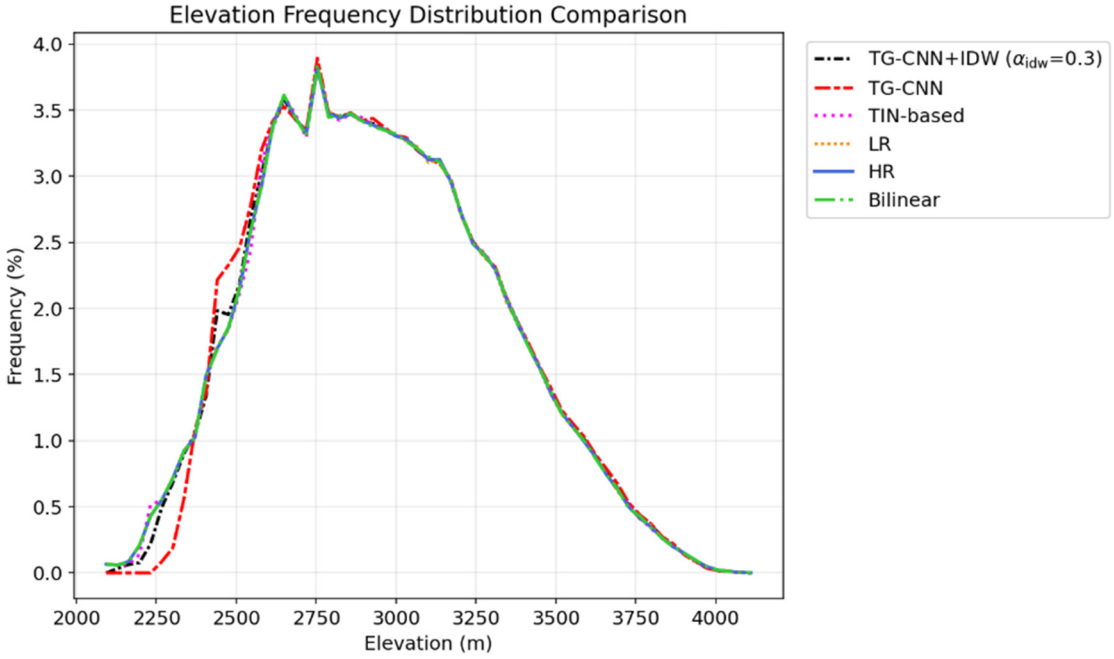


Figure 5. Elevation frequency distribution comparison of DEMs from different methods.

As shown in Table 2 and Figure 5, all the tested reconstruction methods are capable of basically restoring the overall elevation distribution characteristics of real terrain. Nevertheless, the original model, TG-CNN, has a certain degree of deviation in extreme elevation values and partial frequency distribution compared with the HR reference DEM. After being equipped with IDW post-processing, the TG-CNN + IDW scheme exerts a prominent correction effect on the biased elevation results generated by the network. It effectively calibrates the underestimated peak elevation and optimizes the overall elevation frequency distribution.

Taking the real ground-truth low-resolution DEM (LR) as the unified benchmark, this study calculates and compares two core elevation-error metrics, namely, the mean absolute error (MAE) in Equation (13) and root mean square error (RMSE) in (14). The quantitative error calculation covers the prediction results of the TG-CNN model before and after hyper-parameter optimization, TIN-based terrain simplification results, and bilinear resampling results. By quantitatively analyzing the MAE and RMSE values of different

methods, the elevation prediction accuracy and terrain reconstruction performance of the optimized TG-CNN are comprehensively quantified.

$$\mathrm{MAE} = \frac{1}{\mathrm{H} \times \mathrm{W}} \sum_{i=1}^{H} \sum_{j=1}^{W} |I_1(i,j) - I_2(i,j)| \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [I_1(i,j) - I_2(i,j)]^2} \tag{14}$$

Although TG-CNN converges and produces results with reasonable error magnitudes (Table 3), its reconstruction quality is outperformed by conventional resampling and the TIN-based method.

Table 3. Comparison of loss (the 90 m LR DEM is adopted as the reference benchmark for calculation).

Method	MAE	RMSE
TG-CNN	13.7441	25.0843
TG-CNN + IDW(α_{idw} = 0.3)	4.8066	8.0737
Bilinear	1.4073	1.9782
TIN-based	3.3423	5.2397

4.2. Visual Comparison of Hillshade Rendering and Terrain Ridge and Valley Lines

To intuitively demonstrate the terrain shape preservation capability of the TG-CNN model before and after hyper-parameter optimization, the global hillshade rendering of all experimental DEM results was performed using GlobalMapper (Version 21.0). Visual terrain comparison can effectively compensate for the limitations of single quantitative error indicators such as MAE and RMSE, which only reflect overall numerical errors. Through hillshade rendering, the subtle topographic undulations, overall mountain morphology, and local terrain details reconstructed by different methods can be visually displayed.

As can be observed from the hillshade visualization in Figure 6, all comparison methods achieve valid terrain generalization to a certain extent. The vanilla TG-CNN produces a relatively strong smoothing effect during down-scaling, which blurs some subtle microtopographic undulations. After being post-processed by IDW fusion with α_{idw} = 0.3, the TG-CNN + IDW result recovers partial topographic details and local relief features, making the rendered landform closer to the geomorphic texture of the reference LR DEM.

As shown in Figure 7, all tested methods retain the main ridge-and-valley skeleton of the study area. The standalone TG-CNN, however, removes a large number of secondary branch skeleton lines, leaving the local terrain topology incomplete. Introducing IDW post-processing at α_{idw} = 0.3 recovers many of these discarded secondary lines and yields a more complete and continuous skeleton than the standalone model, but this recovery is largely attributable to the dominant IDW component rather than to the network itself.

To quantitatively assess the preservation of terrain structural lines, the ridge and valley lines extracted from the 90 m reference DEM are adopted as the ground-truth benchmark. A unified spatial buffer tolerance (one 90 m cell) is defined for vector-line matching, and precision, recall, F1-score, and the average horizontal offset distance are computed for the reconstruction results: bilinear resampling, TIN-based interpolation, the standalone TG-CNN, and the TG-CNN + IDW fused result at α_{idw} = 0.3.

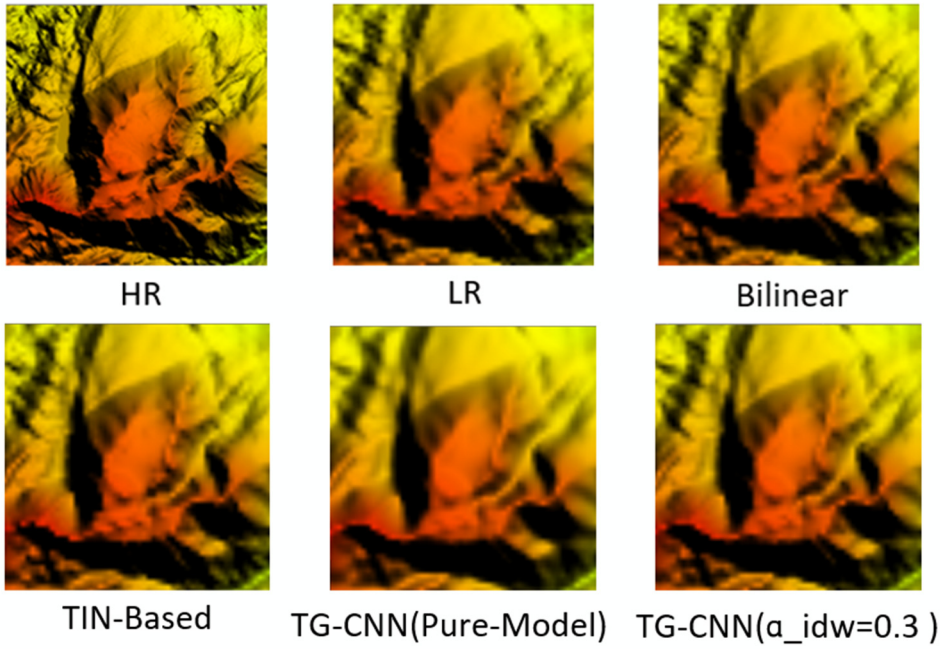


Figure 6. Comparison of terrain generalization results.

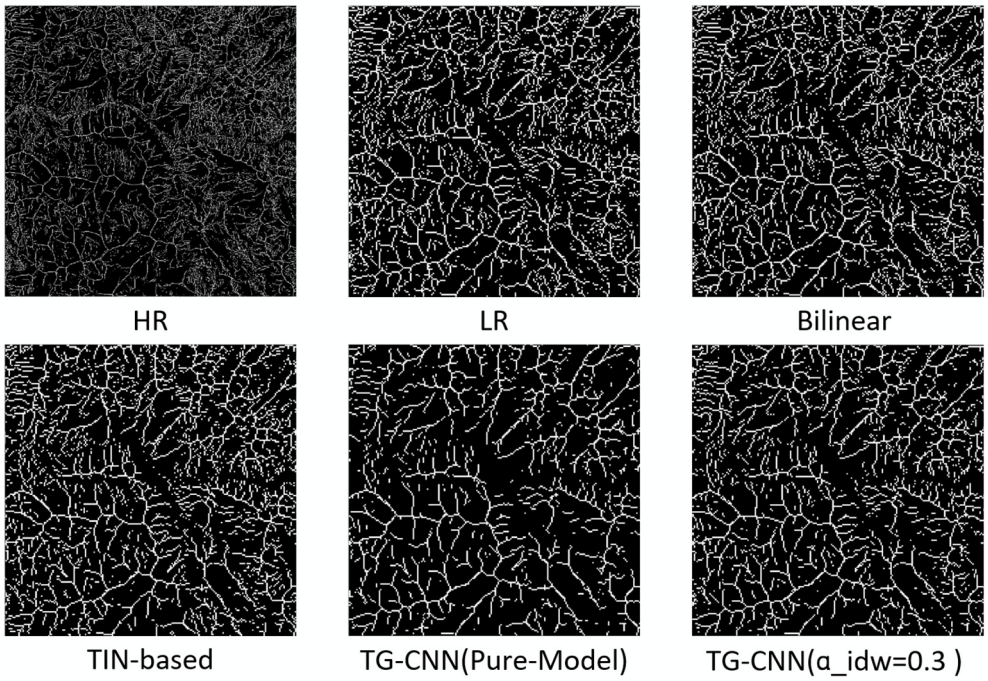


Figure 7. Ridge- and valley-line comparison.

We adopt line-length-based TP, FP and FN for skeleton evaluation:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{15}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16}$$

$$\text{F1} = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

where TP, FP, and FN denote the length of matched terrain lines, falsely extracted lines, and missing reference lines within pre-defined spatial buffer tolerance, respectively.

The average horizontal offset quantifies the geometric deviation between predicted and reference ridge–valley vectors:

$$\mathrm{MHO} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \sqrt{(\mathrm{x}_{\mathrm{ref,i}} - \mathrm{x}_{\mathrm{pred,i}})^2 + (\mathrm{y}_{\mathrm{ref,i}} - \mathrm{y}_{\mathrm{pred,i}})^2} \quad (18)$$

in which $(x_{ref,i}, y_{ref,i})$ are the coordinates of sampled points on ground-truth skeleton lines and $(x_{pred,i}, y_{pred,i})$ are the nearest corresponding points on extracted lines.

Table 4 compares ridge-line extraction against the 90 m reference ridge mask, with all masks being registered to the common 90 m grid. Bilinear and TIN attain the highest F1-scores (0.944 and 0.941), followed by the TG-CNN + IDW fusion model (0.933) and the standalone TG-CNN (0.843). The F1 ranking is driven by recall: the standalone model retains only 73.5% of the reference ridge cells because its smoother surface suppresses roughly a quarter of the true skeleton. In precision, however, the standalone TG-CNN outperforms the interpolation baselines—its precision (0.988) exceeds that of bilinear (0.944) and TIN (0.955), and its average offset (24 m) is smaller than both—indicating that the ridges that it does detect are almost always genuine and well positioned. The network therefore behaves as an aggressive filter, keeping the most certain ridge lines while discarding real secondary branches. The fusion at α = 0.3 attains the highest precision (0.992) and the smallest offset (11 m), largely because its 70% IDW component interpolates directly from the 90 m nodes; blending in the network does not improve on IDW alone. Overall, the network does not enhance ridge preservation: at full model weight, it trades completeness for precision, and its F1 remains below all interpolation baselines.

Table 4. Quantitative evaluation metrics of extracted ridge and valley lines with different methods.

Method	Precision	Recall	F1-Score	Average Horizontal Offset Distance (m)
TG-CNN (pure)	0.9884	0.7345	0.8428	24.18
TG-CNN + IDW (α_{idw} = 0.3)	0.9924	0.8799	0.9328	11.01
Bilinear	0.9438	0.9447	0.9408	36.64
TIN-based	0.9553	0.9268	0.9443	46.90

4.3. Loss Function Coefficient Analysis

To evaluate the effects of slope–aspect coefficients within the error function on model training, three random seeds (s0, s1, and s2) were adopted. With max pooling as the down-sampling operator, experiments were carried out under multiple coefficient combinations: (0.6, 0.4), (0.6, 0), (0, 0.4), (0.3, 0.2), and (0, 0). The Valdez dataset was used for model training, and model performance was assessed on the Chongqing validation dataset to determine the optimal model behavior.

Table 5 reports the validation-set accuracy of the five loss-coefficient settings, with each of the three training seeds being listed separately. The within-configuration seed variation is large—up to 4.9 m in MAE for a single coefficient setting (e.g., 11.55–16.46 m at α = 0.6 and β = 0)—and it exceeds the differences between configurations, whose per-seed means span only 12.95–13.93 m (MaxPool). No weighted setting, including the slope/aspect-regularized losses, systematically improves upon the pure elevation loss (α = β = 0), and the combined setting (α = 0.6 and β = 0.4) yields the largest mean error. Consequently, the coefficient setting adopted for the final evaluation (α = 0.6, β = 0, and seed 1) was chosen on the validation set, but its configuration-level performance is indistinguishable from that

of the other settings, indicating that its favorable single-seed result is a seed effect rather than evidence that the geomorphic loss terms improve generalization.

Table 5. Sensitivity of validation accuracy (Chongqing) to loss-function coefficients α and β, per training seed.

α β	Seed	MaxPool MAE	MaxPool RMSE	AvgPool MAE	AvgPool RMSE
(0.6,0.4)	S0	12.61	25.19	16.12	32.22
	S1	16.28	32.77	12.42	22.64
	S2	12.91	22.79	15.68	28.15
(0.3,0.2)	S0	13.15	23.07	13.61	22.52
	S1	12.13	20.55	15.75	25.97
	S2	13.58	22.72	13.76	23.78
(0.0,0.0)	S0	14.95	24.44	14.34	21.32
	S1	12.03	18.89	15.14	24.54
	S2	13.37	21.31	13.93	20.56
(0.6,0.0)	S0	11.87	19.76	12.16	18.70
	S1	11.55	19.04	14.23	23.94
	S2	16.46	29.17	19.92	37.88
(0.0,0.4)	S0	12.64	21.60	15.30	27.59
	S1	15.75	32.20	15.88	31.01
	S2	13.35	23.55	16.25	28.30

Table 5 also enables a comparison of the two downsampling-pooling variants. At the configuration level, MaxPool achieves a lower mean MAE than AvgPool for all five coefficient settings (gaps of 0.8–2.1 m), with overall means of 13.51 m versus 14.97 m (MAE) and 23.74 m versus 25.94 m (RMSE) across the 15 trained models of each variant. At the seed level, the comparison is noisier: MaxPool wins 13 of the 15 seed pairs, with two exceptions (e.g., $\alpha = 0.6$, $\beta = 0.4$, and seed 1, where AvgPool reaches 12.42 m versus 16.28 m for MaxPool). Nevertheless, the pooling choice exerts a larger and more consistent effect than any loss-coefficient adjustment, and it is consistent with the design intent: max pooling retains the extreme value of each pooling window and hence better preserves the dominant crests and troughs of terrain. The MaxPool variant was accordingly adopted for the final evaluation.

As shown in Table 6, the error grows monotonically with α: pure IDW ($\alpha_{idw} = 0$) achieves the best accuracy (MAE of 1.78 m and RMSE of 2.58 m), and every increase in the network's share degrades the fused result, reaching an MAE of 13.74 m at $\alpha_{idw} = 1$.

Table 6. MAE and RMSE errors for combinations of different IDW fusion methods.

Method	MAE	RMSE
TG-CNN	13.7441	25.0843
$\alpha_{idw} = 0.7$	9.8510	17.7258
$\alpha_{idw} = 0.5$	7.2915	12.8556
$\alpha_{idw} = 0.3$	4.8066	8.0737
IDW	1.7759	2.5833

4.4. IDW Fusion Coefficient Analysis

This section applies the optimal model selected on the validation set (MaxPool, slope weight 0.6, aspect weight 0, and seed 1) to the test region. Its output is then blended with IDW interpolation at five fusion coefficients, $\alpha \in \{0, 0.3, 0.5, 0.7, 1.0\}$, to examine the effect of IDW post-processing, and the MAE and RMSE of each fused result are computed against the 90 m reference DEM.

Shown in Figure 8, all parameter combinations realize effective terrain generalization. Visually, the two pooling approaches produce consistent progressive changes in surface smoothing as the IDW weight α_{idw} varies. A smaller α_{idw} value retains more fine topographic textures and weakens over-smoothing, while the elevation surface becomes increasingly blurred and generalized with the gradual rise in α_{idw}.

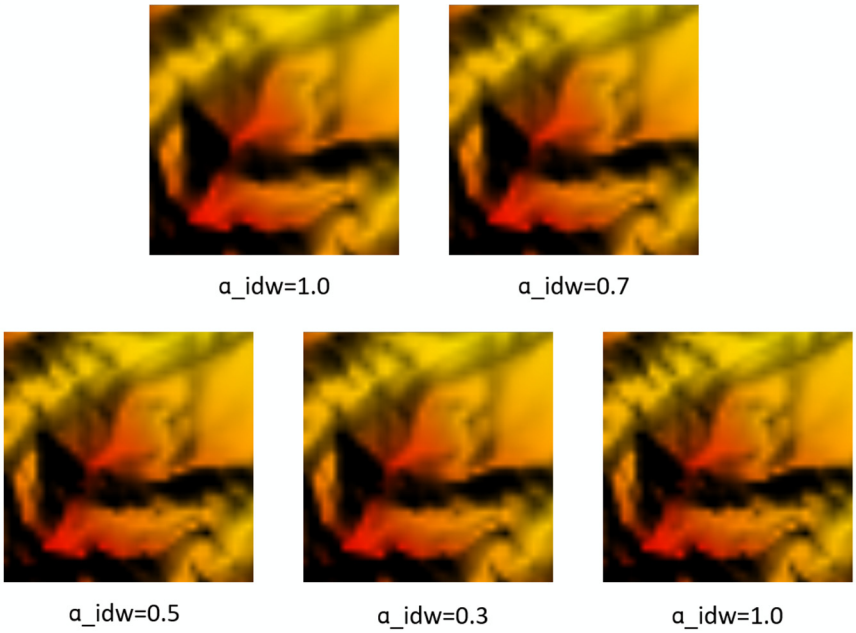


Figure 8. Visualization of results obtained with different downsampling modules and IDW fusion coefficients.

5. Discussion

5.1. Controllable Terrain Generalization Intensity Modulated by IDW Fusion Weight

The IDW fusion weight α provides a direct control over the intensity of terrain generalization. At $\alpha = 0$, the output reduces to the pure IDW interpolation, which is anchored to the original [−1] m sampling nodes and yields the smallest elevation error (MAE of 1.78 m and RMSE of 2.58 m), together with the most complete ridge network (recall ≈ 0.94 for the interpolation reconstructions). As α increases, the network's share grows, and the surface becomes progressively smoother: the number of detected ridge cells declines (from 37,545 at $\alpha = 0.3$ to 29,935 at $\alpha = 1$), and ridge recall falls from 0.88 to 0.73, while precision remains above 0.98. The coefficient therefore behaves as a continuous fidelity–smoothing dial, confirming the mechanism by which IDW fusion modulates generalization intensity.

The corrected results, however, show that this control is currently one-sided. Because the network reconstructs elevation less accurately than the IDW interpolation, every increase in α degrades the fused MAE and RMSE monotonically (from 1.78 m and 2.58 m at $\alpha = 0$ to 13.74 m and 25.08 m at $\alpha = 1$), and no nonzero α improves upon pure IDW. The generalization-intensity control is thus real in mechanism but not yet useful in practice: with the present model the preferred operating point is $\alpha = 0$, and the network currently contributes smoothing without any compensating accuracy benefit. Should the elevation

accuracy of the network be improved in future work, the same fusion framework would allow users to trade fidelity against skeleton-preserving generalization in a principled way, without the penalty observed here.

Figure 9 visualizes how the IDW fusion coefficient shapes the extracted ridge–valley skeleton on the Valdez study area. At the IDW-dominated end, the skeleton follows the reference 90 m DEM closely, retaining the full network of main and secondary branch lines—consistent with the high recall (≈0.94) of the interpolation-based reconstructions (Table 4). As the network's fusion weight increases, the skeleton is progressively refined: the trunk lines persist unchanged while the secondary branches are thinned and eventually removed, until the standalone TG-CNN retains only the most prominent skeleton (recall of 0.73). This refinement does not, however, sacrifice positional reliability. The lines retained by the network are the most precise of all methods: its precision (0.988; 0.992 for the $\alpha = 0.3$ fusion model) exceeds that of bilinear (0.944) and TIN (0.955), and its average horizontal offset (24 m; 11 m for the fusion) is smaller than both baselines. The fusion weight thus acts as a continuous skeleton-refinement control, and within this completeness–precision trade-off, the network's genuine strength lies in the accuracy of the lines that it keeps.

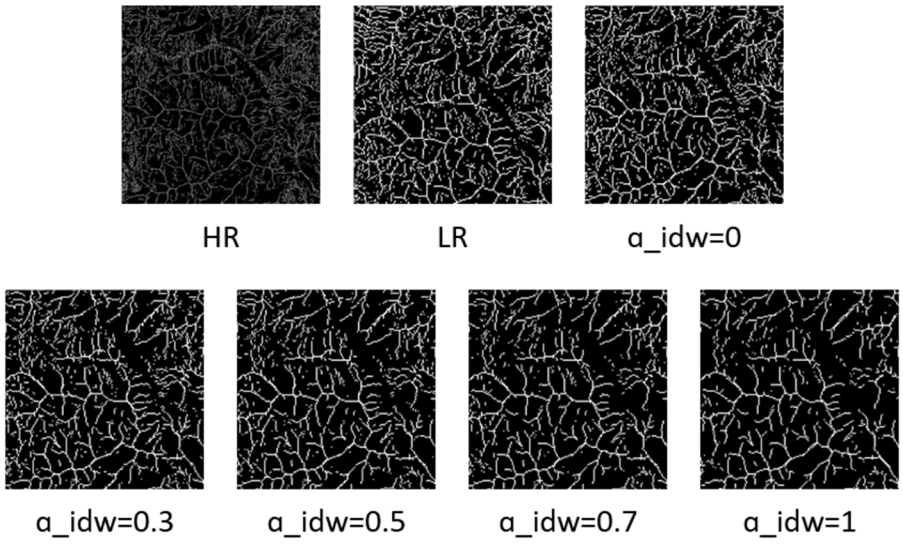


Figure 9. Extracted ridge–valley skeletons under different IDW fusion coefficients (Valdez study area, Maxpool downsampling, and slope–aspect loss weights $\alpha = 0.6$ and $\beta = 0.0$).

5.2. Complementary Effect of Slope and Aspect Geometric Constraints

The slope and aspect terms in the composite loss were designed to complement each other: the slope term penalizes errors in gradient magnitude, while the aspect term—encoded as sine and cosine to preserve circular continuity—penalizes errors in the gradient direction. The intended effect is a surface that not only matches the reference elevations but also inherits the reference terrain's local geometry, thereby suppressing artificial terrain artifacts.

In the corrected experiments, the evidence for this complementary effect is mixed and weak. In the validation region (Table 5), the slope-dominated settings show a mild favorable tendency: the slope-only configuration ($\alpha = 0.6$ and $\beta = 0$)—the configuration of the final model—achieves the best single model (11.55 m) and a mean MAE (13.29 m) slightly below that of the pure elevation loss (13.45 m), as does the moderate setting ($\alpha = 0.3$, $\beta = 0.2$; 12.95 m). The aspect term, by contrast, contributes nothing: both settings with $\beta = 0.4$ perform worse than their $\beta = 0$ counterparts (13.93 m versus 13.29 m; 13.91 m versus 13.45 m), and the combined setting yields the largest mean error of all five configurations. None of these differences, however, exceeds the seed-to-seed variation (up to 2.24 m), so the

favorable tendency of the slope term is suggestive rather than statistically established, and the intended complementarity of the two geometric terms is not observed. Two properties of the task plausibly explain why the effect is so small. First, the elevation loss is already computed on z-score-normalized tiles, so fitting the surface implicitly constrains the local gradient field that the additional terms encode explicitly. Second, the slope and aspect penalties mostly duplicate that signal and add optimization noise. The geometric constraints therefore remain a well-motivated design element—sine–cosine aspect encoding is mathematically sound and avoids the $0^\circ/360^\circ$ discontinuity—and the slope term's tendency is consistent with the design intent, but a measurable complementary benefit has not been demonstrated at the present scale. The one observation most consistent with the intended effect is the high precision of the network's ridge skeleton (0.988; Table 4): the retained lines contain almost no spurious fragments, which is precisely the kind of artifact that the geometric constraints were designed to suppress, although the present ablation cannot separate this effect from the inherent smoothing of the architecture.

5.3. Limitations and Future Extensions

The present study reveals several inherent limitations of deep learning for DEM terrain generalization. The most fundamental is the information bottleneck introduced by aggressive downsampling: the network compresses the fine-resolution input through a stride-3 max-pooling layer before any deep processing, so the majority of the source elevation detail is discarded before learning begins and must be re-synthesized from a reduced representation. Classical interpolators, by contrast, operate directly on the source samples and retain information that the network throws away. This structural asymmetry—rather than insufficient training or hyper-parameter choices—plausibly explains why the corrected model remains far from the interpolation baselines in MAE and RMSE in every study region, and it suggests that a purely learning-based reconstruction will struggle to surpass mature interpolation on pixel-level accuracy as long as its front end discards most of the input.

Deep learning in terrain generalization also faces more general difficulties. Terrain relief varies enormously across regions, and a network trained on one study area transfers its priors only weakly to another, as the three-region chain of this study illustrates. Training is sensitive to stochastic factors—different seeds of an identical configuration produce markedly different models—which weakens reproducibility. Moreover, the learned surface is a statistical reconstruction whose smoothing is uniform and implicit: it cannot distinguish, as a cartographer would, between trivial microtopography that should be filtered and geomorphologically meaningful features that must be preserved, which is exactly why secondary ridge branches were so aggressively removed. Finally, at the modest generalization factor studied here ($3\times$), the task is fidelity-dominated, and classical interpolation is already near-optimal, leaving little room for semantic priors to add value; the potential of learning-based generalization lies instead in larger generalization factors and multi-purpose outputs, where interpolation has no established solution.

Another important practical constraint comes from our dataset coverage: all training, validation, and test sites in this work are mountain-dominated landscapes. Plains and hilly–plain transition zones are not included in our experimental datasets. Accordingly, the transferability, terrain generalization quality, and practical applicability of TG-CNN for non-mountain landforms remain unvalidated and require further dedicated experiments.

For future work, the most promising directions follow directly from these limitations: learning residuals with respect to a classical interpolator rather than the full surface, replacing destructive pooling with information-preserving downsampling, and stabilizing training across seeds and regions.

6. Conclusions

This study addressed the long-standing need for a learning-based terrain generalization method by constructing TG-CNN, a feature-fusion convolutional neural network for 3× DEM down-scaling. The network embeds geomorphic priors—slope and sine–cosine circularly encoded aspect—into both the input channels and a composite loss function, and offers two downsampling variants (max pooling and average pooling). A leave-one-area-out scheme trained the model on Valdez, validated it on Chongqing, and tested it on the unseen Gore Range, with an adjustable IDW fusion module appended for post-processing.

The experiments lead to three main conclusions. First, in terms of pixel-level elevation accuracy, classical interpolation remains superior: on the test region, bilinear resampling attains an MAE of 1.41 m and an IDW of 1.78 m, whereas the standalone TG-CNN attains 13.74 m. The gap persists across all three study regions and across thirty configurations and seeds, indicating a structural information bottleneck of the pooling-based front end rather than insufficient training. Second, the IDW fusion weight provides a continuous control over generalization intensity: as the network's share increases, the surface is progressively smoothed and the ridge–valley skeleton is progressively refined, from a complete skeleton with all secondary branches to a sparse but highly reliable trunk network (precision of 0.988 and offset of 24 m, versus 0.944 and 37 m for bilinear). This is a genuine scale-driven abstraction effect rather than mere elevation fitting. However, because every increase in the network's share degrades the fused elevation error monotonically, the currently useful operating point of this control is the pure IDW end, and the fusion benefit observed at $\alpha_{idw} = 0.3$ derives almost entirely from the IDW component. Third, the sensitivity analyses show that the max-pooling variant consistently outperforms average pooling, while the loss-coefficient settings are statistically indistinguishable and dominated by seed variation, so the slope/aspect regularization, though well-motivated, has not yet produced a measurable accuracy benefit.

In summary, TG-CNN demonstrates that deep networks can perform controllable, structure-aware terrain abstraction, but deep learning does not yet surpass mature interpolation in terms of elevation accuracy at this generalization factor. Future work should pursue residual learning with respect to a classical interpolator, information-preserving downsampling architectures, and multi-region training, which directly target the limitations identified here.

Author Contributions: Conceptualization, Y.W. (Yanchen Wan), Y.W. (Yue Wang) and W.J.; methodology, Y.W. (Yanchen Wan), W.J. and J.F.; software, Y.W. (Yanchen Wan), X.L., J.F. and H.J. (Haowen Jiang); validation, Y.W. (Yanchen Wan) and H.J. (Han Jiang); formal analysis, Y.W. (Yanchen Wan) and Y.W. (Yue Wang); resources, W.J.; data curation, Y.W. (Yanchen Wan), H.J. (Han Jiang) and X.L.; writing—original draft preparation, Y.W. (Yanchen Wan) and X.L.; writing—review and editing, W.J., Y.W. (Yue Wang), H.J. (Han Jiang) and H.J. (Haowen Jiang); visualization, Y.W. (Yanchen Wan) and X.L.; supervision, W.J.; project administration, W.J.; funding acquisition, W.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research study was financially supported by the National Key Research and Development Program of China [grant number 2022YFB3904100], the National Natural Science Foundation of China [grants number 42571515 and 41371428], and LIESMARS Special Research Funding.

Data Availability Statement: The training and test digital elevation model (DEM) data used in this study were partially derived from the SRTM dataset (available at https://www.earthdata.nasa.gov/data/instruments/srtm, accessed on 13 January 2026.) and the remaining part from the sample DEM dataset at https://shadedrelief.com/SampleElevationModels/ (accessed on 13 January 2026). The code, figures and DEMs that supports the findings of this study are openly available in Figshare at https://doi.org/10.6084/m9.figshare.31437316 (accessed on 1 March 2026).

Acknowledgments: The authors gratefully acknowledge all data providers whose data play a crucial role in the experiment in this article. Special thanks are extended to the anonymous reviewers and the editors of the journal for their important contributions in improving the article for publication.

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

References

1. Ai, T.; Li, J. A DEM generalization by minor valley branch detection and grid filling. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 198–207. [CrossRef]
2. Samsonov, T. Granularity of digital elevation model and optimal level of detail in small-scale cartographic relief presentation. *Remote Sens.* **2022**, *14*, 1270. [CrossRef]
3. Weibel, R. Models and experiments for adaptive computer-assisted terrain generalization. *Cartogr. Geogr. Inf. Syst.* **1992**, *19*, 133–153. [CrossRef]
4. Xia, Y.; Ho, A.T.S.; Ji, Y. A novel wavelet stereo matching method to improve DEM accuracy generated from SPOT stereo image pairs. In Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium, Sydney, Australia, 9–13 July 2001; Volume 7, pp. 3277–3279. [CrossRef]
5. Jenny, B.; Jenny, H.; Hurni, L. Terrain generalization with multi-scale pyramids constrained by curvature. *Cartogr. Geogr. Inf. Sci.* **2011**, *38*, 110–116. [CrossRef]
6. Ma, T.; Chen, Y.; Hua, Y.; Chen, Z.; Chen, X.; Lin, C.; Yang, C. DEM generalization with profile simplification in four directions. *Earth Sci. Inform.* **2017**, *10*, 29–39. [CrossRef]
7. Chen, Z.-T.; Guevara, J.A. Systematic selection of very important points (VIP) from digital terrain models for constructing triangular irregular networks. In Proceedings of the Auto-Carto 8: Eighth International Symposium on Computer-Assisted Cartography, Baltimore, MD, USA, 29 March–3 April 1987; pp. 50–56.
8. Fei, L.; He, J.; Ma, C. Three-dimensional Douglas-Peucker algorithm and the study on its application to automated generalization of DEM. In Proceedings of the GeoInformatics 2006: Geospatial Information Science, Wuhan, China, 28–29 October 2006; Volume 6420, p. 64200G. [CrossRef]
9. Heckbert, P.S.; Garland, M. Survey of polygonal surface simplification algorithms. In Proceedings of the SIGGRAPH '97 Course Notes: Multiresolution Surface Modeling, Los Angeles, CA, USA, 5 August 1997.
10. Li, Z. An algorithm for compressing digital contour data. *Cartogr. J.* **1988**, *25*, 143–146. [CrossRef]
11. Ai, T. The drainage network extraction from contour lines for contour line generalization. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 93–103. [CrossRef]
12. Li, J.; Ai, T. The DEM generalization based on the simplification of valley system. In *Proceedings of the GeoInformatics 2007: Geospatial Information Science, Nanjing, China, 25–27 May 2007*; Chen, J., Pu, Y., Eds.; SPIE: Bellingham, WA, USA, 2007; Volume 6753, p. 675328.
13. Palomar-Vázquez, J.; Pardo-Pascual, J. Automated spot heights generalisation in trail maps. *Int. J. Geogr. Inf. Sci.* **2008**, *22*, 91–110. [CrossRef]
14. Zhao, B.; Zhu, Z.; Liu, G. Extraction of terrain feature points and terrain branches based on contour curve geometric structure subdivision using a constrained Delaunay triangular network. *J. Geo-Inf. Sci.* **2025**, *27*, 2422–2439. [CrossRef]
15. Li, J.; Ai, T.; Wang, H. The DEM generalization based on the filling valley coverage. *Acta Geod. Cartogr. Sin.* **2009**, *38*, 272–275.
16. Li, J.; Shen, Y.; Gao, A.; Chen, K.; Zhang, Y.; Ai, T. A heuristic DEM generalization by combining catchments. *Geocarto Int.* **2022**, *37*, 9392–9407. [CrossRef]
17. Wu, Q.; Chen, Y.; Wilson, J.P.; Liu, X.; Li, H. An effective parallelization algorithm for DEM generalization based on CUDA. *Environ. Model. Softw.* **2019**, *114*, 64–74. [CrossRef]
18. Chen, Y.; Ma, T.; Chen, X.; Chen, Z.; Yang, C.; Lin, C.; Shan, L. A new DEM generalization method based on watershed and tree structure. *PLoS ONE* **2016**, *11*, e0159798. [CrossRef] [PubMed]
19. Chen, Z.; Ma, X.; Yu, W.; Wu, L. An integrated graph Laplacian downsample (IGLD)-based method for DEM generalization. *Earth Sci. Inform.* **2020**, *13*, 973–987. [CrossRef]
20. Zhao, J.; Yuan, Y.; Dong, Y.; Li, Y.; Shao, C.; Yang, H. Void filling of digital elevation models based on terrain feature-guided diffusion model. *Remote Sens. Environ.* **2024**, *315*, 114432. [CrossRef]
21. Bittner, K.; Zorzi, S.; Krauß, T.; d'Angelo, P. DSM2DTM: An end-to-end deep learning approach for digital terrain model generation. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *X-1/W1-2023*, 925–933. [CrossRef]
22. Yang, J.; Xu, J.; Lv, Y.; Zhou, C.; Zhu, Y.; Cheng, W. Deep learning-based automated terrain classification using high-resolution DEM data. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103249. [CrossRef]
23. Ho, Y.-F.; Grohmann, C.H.; Lindsay, J.; Reuter, H.I.; Parente, L.; Witjes, M.; Hengl, T. GEDTM30: Global ensemble digital terrain model at 30 m and derived multiscale terrain variables. *PeerJ* **2025**, *13*, e19673. [CrossRef] [PubMed]

24. Syzdykbayev, M.; Karimi, B.; Karimi, H.A. A method for extracting some key terrain features from shaded relief of digital terrain models. *Remote Sens.* **2020**, *12*, 2809. [CrossRef]
25. Zheng, X.; Bao, Z.; Yin, Q. Terrain self-similarity-based transformer for generating super-resolution DEMs. *Remote Sens.* **2023**, *15*, 1954. [CrossRef]
26. Chen, Z.; Wang, X.; Xu, Z.; Hou, W. Convolutional neural network based DEM super resolution. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *XLI-B3*, 247–250. [CrossRef]
27. Li, M.; Dai, W.; Wang, G.; Wang, B.; Chen, K.; Gao, Y.; Obiri Yeboah Amankwah, S. Reconstructing high-resolution DEMs from 3D terrain features using conditional generative adversarial networks. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *133*, 104115. [CrossRef]
28. Zhou, A.; Chen, Y.; Wilson, J.P.; Chen, G.; Min, W.; Xu, R. A multi-terrain feature-based deep convolutional neural network for constructing super-resolution DEMs. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *120*, 103338. [CrossRef]
29. 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]
30. Gao, B.; Yue, L.; Qiu, Z.; Yuan, Q.; Shen, H. Exploring the integration of auxiliary modal information from remote sensing images for DEM super-resolution. *Int. J. Geogr. Inf. Sci.* **2025**, *39*, 2886–2912. [CrossRef]
31. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
32. Kennelly, P.J.; Patterson, T.; Jenny, B.; Huffman, D.P.; Marston, B.E.; Bell, S.; Tait, A.M. Elevation models for reproducible evaluation of terrain representation. *Cartogr. Geogr. Inf. Sci.* **2021**, *48*, 63–77. [CrossRef]
33. Weibel, R. Generalization of spatial data: Principles and selected algorithms. In *Algorithmic Foundations of Geographic Information Systems*; van Kreveld, M., Nievergelt, J., Roos, T., Widmayer, P., Eds.; Springer: Berlin/Heidelberg, Germany, 1997; pp. 99–152. [CrossRef] [PubMed]
34. Lindsay, J.B. Whitebox GAT: A case study in geomorphometric analysis. *Comput. Geosci.* **2016**, *95*, 75–84. [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.