

# From gaps to granularity: CRPAG-DSHAT based multi-modal deep learning framework for DEM void repair and super-resolution reconstruction in Himalayas

Sayantan Mandal <sup>\*</sup>, Ashis Kumar Saha

*Department of Geography, Delhi School of Economics, University of Delhi, New Delhi, 110007, India*

## ARTICLE INFO

**Keywords:**  
 Deep neural networks (DNNs)  
 Void filling  
 Super resolution  
 Multi-modal data fusion  
 ALOS PALSAR DEM

## ABSTRACT

Digital Elevation Models (DEMs) are essential for terrain characterization and environmental modeling, yet their utility is limited by data voids and coarse resolution, especially in complex mountainous regions of Himalayas. To address these challenges, we propose a novel dual-stage deep learning pipeline that unifies void filling and super-resolution into a cohesive framework, leveraging both topographic fidelity and spectral texture. First, the **Conditional Residual Pyramid Attentional Generator (CRPAG)** a hybrid model that integrates multi-scale DEM features with Sentinel-2 red band reflectance (~665 nm) using an **Improved Channel Attention Module (ICAM)**, **Residual Pyramid Attention Block (TFG\_RPAB)**, and a dual-encoder design. This allows CRPAG to prioritize structural fidelity (RMSE 9.1–28.9 m) while reconstructing missing terrain features (Mean Absolute Error MAE 1.9–8.1 m). This void-filled, high-resolution DEM then supervises the training of **Dual-Stream Hierarchical Attention Transformer (DS-HAT)**, which performs super-resolution on globally available low-resolution DEMs (ALOS PALSAR), guided by pixel-wise height attention and texture-aware mechanisms. Compared to benchmark models such as MCUNet-EDF and conventional U-Nets, our integrated system shows improvements in elevation accuracy (RMSE ↓, P95 = 9.2 m), spatial consistency (Moran's I ↑), and structural similarity (SSIM ↑), particularly across high-curvature and spectrally ambiguous regions. Besides, Ablation studies confirm the complementary applications of topographic variables in mitigating oversmoothing and enhancing terrain realism. This dual-stage strategy not only enhances DEM fidelity but also provides a scalable framework for improving DEM quality. Through this multi-modal fusion, this work transforms topographic knowledge into computable framework, advancing DEM applicability in hydrological modeling, detection mechanisms and disaster prediction.

## 1. Introduction

Digital Elevation Models (DEMs) are fundamental geospatial datasets that provide a digital representation of the Earth's surface (Zhao et al., 2024). They serve as the essential backbone in a wide spectrum of applications, from hydrological analysis and landslide detection to global sea level monitoring and climate change studies (Yang et al., 2011; Zhang et al., 2024). By conveying topographic and geomorphological information, DEMs enable researchers and practitioners to extract key terrain features and simulate complex geophysical processes (Xiong et al., 2024; Zandsalimi et al., 2024). Besides, precise high resolution DEMs are indispensable in performing detailed ecological risk assessments and monitoring the impacts of climate change, thereby

influencing urban planning and disaster management strategies.

A high-quality DEM is essential not only for enhancing spatial modeling accuracy but also for ensuring process fidelity in simulations. Numerous studies have shown that the spatial resolution and vertical accuracy of DEMs directly influence the performance of downstream geospatial models, particularly in rugged or hydrologically sensitive terrains (Xia et al., 2023). Despite their wide applicability, DEMs are not without limitations. DEMs derived from satellite altimetry, radar interferometry, or photogrammetry often suffer from voids, especially in areas with dense vegetation, steep slopes, shadows, or persistent cloud cover (Dong et al., 2020). These missing pixels disrupt the continuity of terrain representation, leading to significant degradation in the reliability of terrain-based modeling tasks (Boulton and Stokes, 2018). As

\* Corresponding author.

E-mail addresses: [sayantanfire@gmail.com](mailto:sayantanfire@gmail.com) (S. Mandal), [aksaha@geography.du.ac.in](mailto:aksaha@geography.du.ac.in) (A.K. Saha).

such, addressing these voids and improving the resolution and quality of DEMs like ALOS, SRTM, ASTER and HMA (High Mountain Asia) has emerged as a critical challenge and opportunity in modern geospatial science (Han et al., 2025; Luedeling et al., 2007; Zhang et al., 2025).

Traditional solutions, including manual in-situ surveys and the fusion of auxiliary DEM sources such as GMTED2010 or ASTER GDEMv2 are constrained by high labor costs, non-uniform data quality, and limited global availability (Karkee et al., 2008; Li et al., 2020; Milan et al., 2011). In response to this, Interpolation-based methods such as kriging, spline, and inverse distance weighting (IDW) have seen widespread adoption (Reuter and Jarvis, 2007), yet often fail in complex terrain where local methods cannot preserve global topographic fidelity or hydrologic coherence (Arun, 2013; Heritage et al., 2009). Very recently Deep learning-based techniques have emerged as transformative tools for void reconstruction (Gavrili et al., 2019; Jiang et al., 2023; Yuan et al., 2020). Architectures such as convolutional neural networks (CNNs) and conditional generative adversarial networks (CGANs) have demonstrated remarkable capacity to synthesize terrain information by learning both local features and global contextual dependencies (Qiu et al., 2019; Zhao et al., 2024).

However, despite the improved completeness of such reconstructions, void-filled DEMs alone do not inherently guarantee preservation of natural topographic continuity, especially in hydrologically sensitive contexts (Zhang et al., 2024). In this scenario, over-smoothed and under-smoothed void reconstructions in valleys and ridges propagate errors in downstream applications like hydrological modeling by modifying flow paths, generating artificial depressions, or bias slope-dependent indices. In such contexts, DEM Super-Resolution (SR) emerges as a powerful data-centric solution. Unlike conventional upsampling and interpolation (Arun, 2013), SR-DEM frameworks aim to reconstruct terrain features at finer spatial scales by learning complex, nonlinear mappings from low-resolution (LR) input to high-resolution (HR) output, often informed by auxiliary data sources (Han et al., 2025). For example, recent studies have employed deep residual networks and normalizing flow models to capture the conditional distribution of high-resolution DEMs, thereby reducing artifacts such as oversmoothing and preserving critical topographic details (Yu et al., 2025). Similarly, deep learning approaches based on convolutional neural networks and adversarial training (Jiao et al., 2020) underscore the advantages of learning high-order spatial features over conventional interpolation. DEM SR often faces unique challenges compared to natural images, as elevation accuracy must align with geophysical laws. While CNNs (e.g., SRCNN) and transformers (e.g., SwinIR) dominate SR (Cheng et al., 2022), their application to DEMs often neglects auxiliary topographic variables (slope, curvature) that constrain plausible reconstructions. To solve this, (Wang et al., 2024), unified SR and segmentation for remote sensing imagery, highlighting the value of multi-task architectures, which is a concept yet to be adapted for DEM-specific tasks. Hybrid losses combining Mean Absolute Error (MAE) with gradient penalties have improved slope preservation but lack integration with void-filled intermediates (Cao and Weng, 2024).

The choice of DEM source is also a very important aspect as it impacts the terrain analysis outcomes (Xu et al., 2021). We observed that, high-resolution DEMs such as the 8 m HMA DEM are available for selective mountainous regions, their global applicability remains limited both spatially and functionally due to pervasive data voids and regional coverage constraints (Liu et al., 2019). In contrast, global DEMs like ALOS PALSAR (12.5 m) offer wide availability and consistent baseline quality (Costa et al., 2021), making them ideal for scalable elevation enhancement pipelines. However, ALOS's resolution often under-represents critical terrain features in complex regions, particularly in hydrologically sensitive basins, where terrain continuity is essential for accurate process modeling.

Building on this progress, our study introduces a **dual-stage framework** that intelligently couples **void filling and super-resolution (SR)** to generate void-free, high-fidelity DEMs suitable for

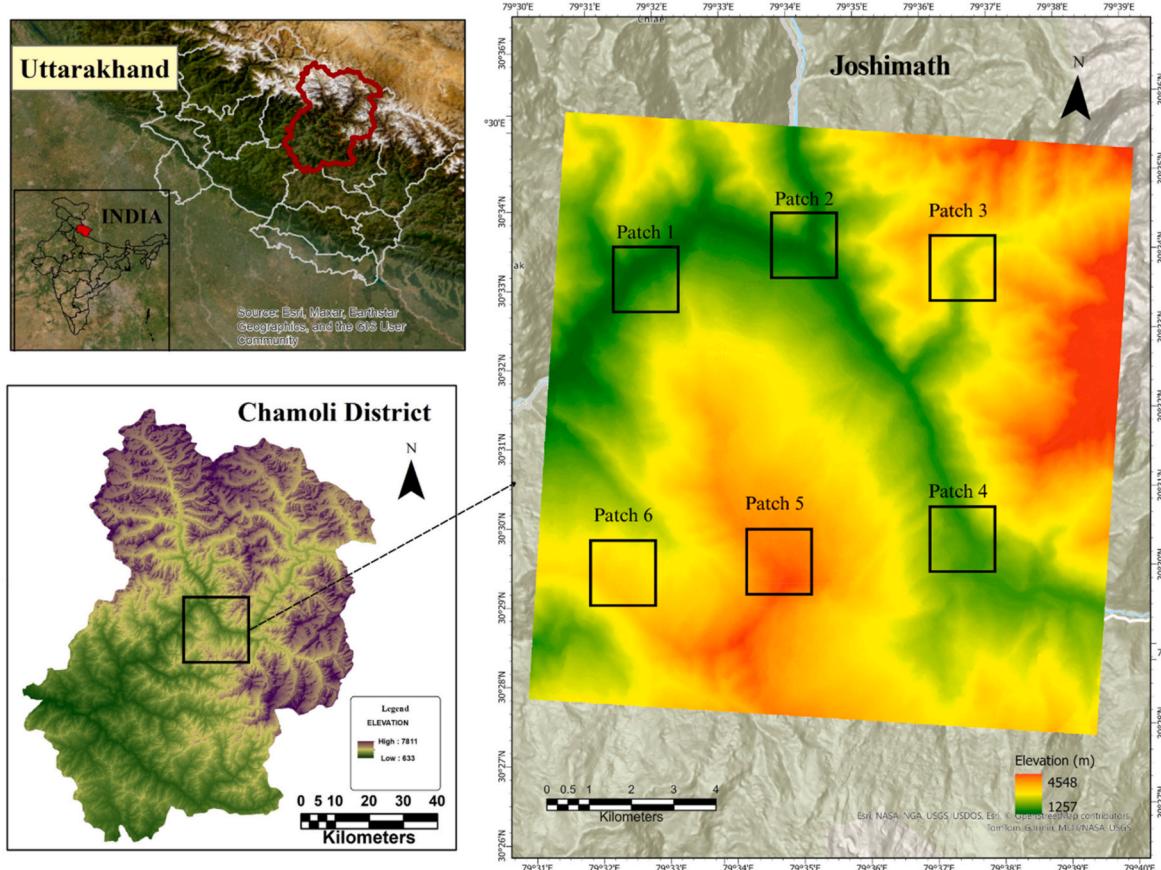
geospatial modeling. First, we reconstruct a void-free HMA DEM using a deep learning-based **Conditional Residual Pyramid Attentional Generator (CRPAG)** that fuses spectral and topographic cues from Sentinel-2 and curvature inputs. While this filled HR DEM could, in theory, serve directly as a modeling input, its spatial coverage is limited to the Himalayan region and may still contain localized smoothing artifacts near filled voids. It is a common issue that propagate hydrological inaccuracies (Boulton and Stokes, 2018). To generalize this terrain fidelity enhancement beyond the HMA region, we leverage the CRPAG-filled 8 m DEM as a supervisory ground truth to train a multi-source SR model **Dual-Stream Hierarchical Attention Transformer (DS-HAT)** that learns to upscale globally available ALOS DEMs. This approach of generating SR DEM enables the creation of high-resolution surrogates from widely available low-resolution sources while inheriting the realism of filled high resolution (HR) terrain. Crucially, the DS-HAT architecture integrates terrain derivatives (e.g., slope, curvature, stream networks) and spectral information (Sentinel-2 Bands) to retain topographic continuity (Clerici et al., 2017) across void-prone and complex terrain zones, mitigating the smoothing limitations of standalone void-filling techniques.

By treating void-filling not as an endpoint but as an **intermediate supervisory bridge** to super-resolution learning, our pipeline ensures that terrain surfaces reconstructed from low-res open source DEMs not only match the statistical fidelity of HR references but also preserve critical landform transitions essential for hydrological and geomorphological modeling. This is critical for applications like hydrological modeling, landslide detection or glacier monitoring, where consistent, high-fidelity DEMs are needed. Thus, by integrating open-source ALOS data with regionally optimized HMA references, this work exemplifies how localized AI training can yield globally scalable solutions, which is an effective shift from prior studies focused on isolated void filling or SR DEM generations. In doing so, we advance a robust, globally transferable elevation enhancement paradigm that bridges the spatial limitations of region-specific HR DEMs with the generalizability of low-res global open-source datasets.

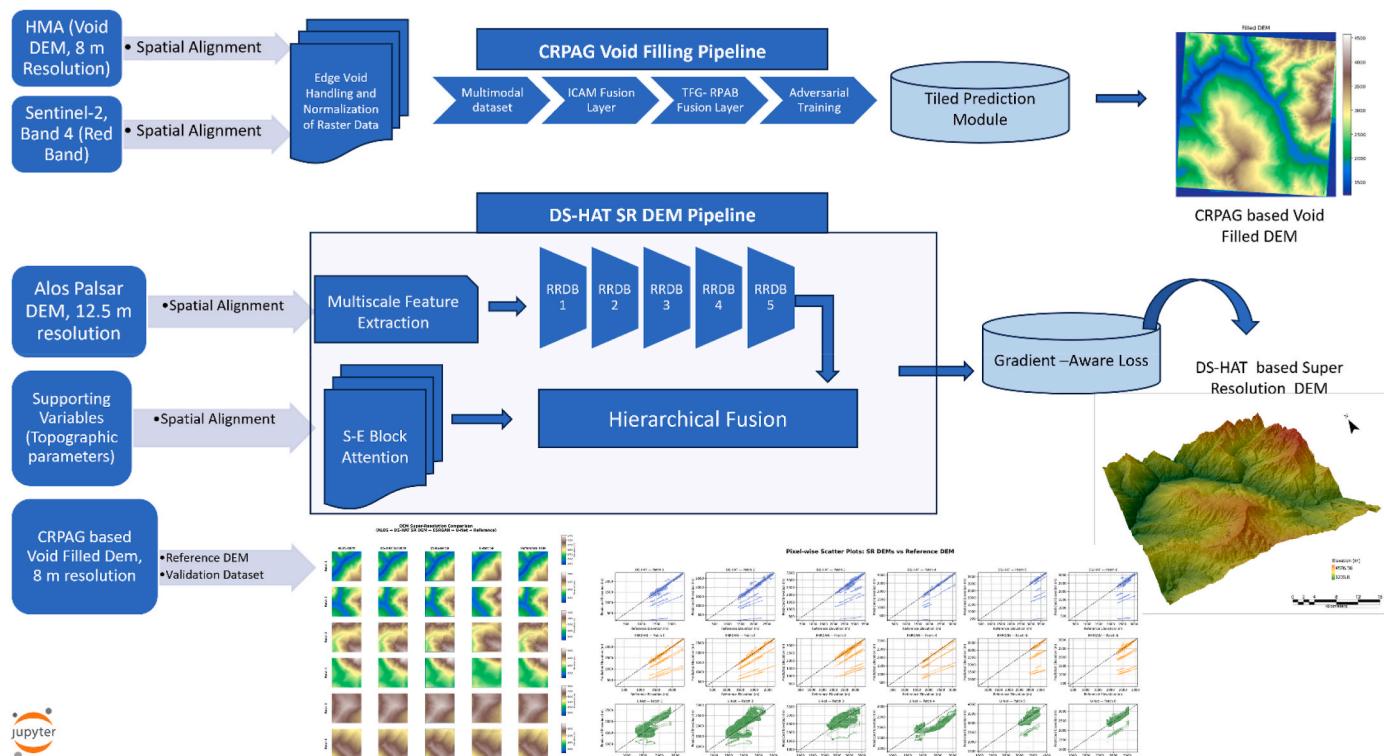
## 2. Study area

The study was conducted over the Joshimath region in Chamoli District ( $30.57^{\circ}\text{N}$ ,  $79.57^{\circ}\text{E}$ ), located in the Uttarakhand state of the Indian Himalayas (Fig. 1). Situated in the Garhwal Himalayas, the area spans elevations from 1800 m in river valleys to over 7800 m at peaks like Nanda Devi, encompassing a dynamic mixture of high-relief mountains, intermontane basins, and fluvial plains. This geomorphic diversity includes steep slopes ( $>45^{\circ}$ ), deep V-shaped valleys carved by the Alaknanda and Dhauliganga rivers, glacial cirques, and exposed bedrock cliffs, features that challenge conventional DEM processing workflows. This area is characterized by rugged mountainous terrain, exhibiting a high degree of topographic heterogeneity, which makes it an ideal study for digital elevation model (DEM) enhancement tasks. The region's rugged terrain, shaped by active tectonic uplift and erosion, generates frequent data voids in optical DEMs due to radar shadowing, cloud cover, and snow persistence. Seasonal snowpack (November–April) and monsoon-clouded summers further exacerbate data gaps, making it a critical testbed for robust void-filling algorithms. Simultaneously, the co-occurrence of high-resolution reference DEMs (e.g., 8 m HMA) and globally available low-resolution datasets (e.g., 30 m ALOS PALSAR) enables controlled super-resolution experiments. The study area is wisely chosen as this area presents a complex geomorphic landscape with extreme topographic heterogeneity and high-relief terrain. Topography, ranging from deep glacial valleys to high-curvature ridgelines, amplifies terrain-induced data discontinuities in conventional DEM products.

Ecologically, Joshimath's land cover ranges from subtropical forests to alpine glaciers, introducing spectral-textural variability that tests multi-modal fusion frameworks making the region a rigorous testbed for



**Fig. 1.** Detailed study area map with the respective patches used on model training for void filling as well as super resolution DEM reconstruction.



**Fig. 2.** Detailed Methodological framework for High topographic fidelity in DEM reconstruction.

both void filling and super-resolution reconstruction. The presence of high-curvature ridges, saddles, and landslide-prone slopes (e.g., 2023 subsidence zones) provides diversified samples for evaluating terrain fidelity in reconstructed DEMs. This area's dual role as a pilgrimage gateway (Badrinath, Hem Kund Sahib) and climate-sensitive zone underscores the societal relevance of high-accuracy elevation data for hazard mitigation and infrastructure resilience. Thus, this study area provides ideal conditions for validating advanced GeoAI-based DEM enhancement frameworks.

### 3. Data and methodology

#### 3.1. Dataset and data pre-processing

This entire study, that has been divided into 2 parts (Fig. 2), utilized a multi-source dataset combining elevation and spectral information to support both DEM void filling and super-resolution tasks in the high-relief region of Joshimath, Uttarakhand, Himalayas. For void filling, a partially incomplete high-resolution DEM (HMA DEM, 8 m) was used as the target for reconstruction. Sentinel-2 Level-2A red band imagery (~10.5 m spatial resolution) was incorporated as an auxiliary input to capture spectral cues related to terrain texture, vegetation, snow, and shadow patterns. A void-free ALOS PALSAR DEM (12.5 m) served as the reference dataset for validation.

For super-resolution DEM generation, the upscaling task was performed on low-resolution ALOS DEMs, using the previously void-filled HMA DEM as the supervisory ground truth. A suite of auxiliary terrain variables was derived from the resampled DEMs to enrich model training, including slope, plan/profile/tangential curvature, flow accumulation, and stream network data (Table 1). Additionally, Sentinel-2 bands B2 (blue), B3 (green), B4 (red), and B8 (NIR) were normalized and fused into the model through a dedicated auxiliary channel.

**Table 1**  
Detailed information of the DEMs and the Auxiliary variables.

Dataset/Variable	Spatial Resolution	Source/Provider	Purpose	Notes/Application	Skewness (Raw → Processed)
<b>ALOS PALSAR DEM</b>	12.5 m	NASA Earth Data	Both	Serves as the primary low-resolution elevation input: <ul style="list-style-type: none"><li>• Void Filling via CRPAG.</li><li>• Super-Resolution via DS-HAT.</li><li>• Provides widely available baseline for scalable DEM enhancement.</li></ul>	0.193 → 0.193
<b>HMA 8 m DEM (High Mountain Asia DEM)</b>	8 m	NSIDC (NASA NSIDC, Version 1)	Both	<ul style="list-style-type: none"><li>• High-resolution ground truth for void-filling (CRPAG) and super-resolution (DS-HAT) training &amp; evaluation.</li><li>• Covers glacier and snow-affected regions of High Mountain Asia.</li><li>• Generated from stereoscopic imagery (GeoEye-1, QuickBird-2, WorldView series).</li></ul>	–
<b>Slope (°)</b>	12.5 m (derived)	Derived from ALOS (ArcGIS Pro)	Super-Resolution	Guides gradient-preserving SR: <ul style="list-style-type: none"><li>• Computed via Sobel kernels.</li><li>• Normalized to [0,1] to prevent scale bias.</li><li>• Helps DS-HAT enforce hydrologically plausible slopes.</li></ul>	0.176 → 0.176
<b>Profile Curvature (rad<sup>-1</sup>)</b>	12.5 m (derived)	Derived from ALOS	Super-Resolution	Captures convex/concave landform curvature: <ul style="list-style-type: none"><li>• Scaled to [0,1] for model stability.</li><li>• Informs DS-HAT attention on flow accumulation zones vs. ridge crests.</li></ul>	0.010 → 0.020
<b>Plan Curvature (rad<sup>-1</sup>)</b>	12.5 m (derived)	Derived from ALOS	Super-Resolution	Encodes lateral curvature of terrain: <ul style="list-style-type: none"><li>• Min-max normalized.</li><li>• Helps DS-HAT maintain correct channel widths and valley shapes during upscaling.</li></ul>	–0.011 → –0.008
<b>Tangential Curvature (rad<sup>-1</sup>)</b>	12.5 m (derived)	Derived from ALOS	Super-Resolution	Measures curvature perpendicular to profile: <ul style="list-style-type: none"><li>• Normalized to [0,1].</li><li>• Used to refine ridge and gully details in DS-HAT outputs.</li></ul>	–0.004 → –0.007
<b>Flow Accumulation</b>	12.5 m (derived)	Derived via D8 algorithm	Both	Indicates drainage density and watershed patterns: <ul style="list-style-type: none"><li>• Log-transformed &amp; normalized.</li><li>• Aids CRPAG in preserving hydrologic continuity.</li><li>• Guides DS-HAT to maintain correct flow network topology post-SR.</li></ul>	1.708 → 1.708
<b>Stream Network</b>	12.5 m (derived)	ALOS hydrologic processing (ArcGIS Pro)	Void Filling	Binary mask of watercourses: <ul style="list-style-type: none"><li>• Ensures CRPAG void fills respect natural channels.</li><li>• Prevents artificial obstruction or removal of streams during inpainting.</li></ul>	7.520 → 7.520
<b>Sentinel-2 Multispectral (MSI)</b>	10–20 m	Copernicus Open Access Hub (ESA)	Void Filling	Provides spectral texture to guide void reconstruction: <ul style="list-style-type: none"><li>• Bands sensitive to vegetation, soils, snow.</li><li>• Co-registered &amp; resampled to DEM grid.</li><li>• Fused via ICAM to inform CRPAG attention weights.</li></ul>	–
<b>Red Band (B4)</b>	10 m	Sentinel-2 MSI	Void Filling	Sensitive to soil and reddened surfaces: <ul style="list-style-type: none"><li>• Key input to ICAM for detecting bare rock and soil patterns.</li><li>• Helps fill voids in shadow-prone high-relief areas.</li></ul>	2.719 → 2.719
<b>HMA Void-Free DEM (TanDEM-X)</b>	12 m	TUM/GFZ/NASA HMA Repository	Reference	High-fidelity ground truth in Himalayas: <ul style="list-style-type: none"><li>• Used to compute CRPAG/DS-HAT loss metrics (RMSE, SSIM, P95).</li><li>• Ensures void filling and SR results are benchmarked against the best available DEM.</li></ul>	–

### 3.2. DEM void filling

#### 3.2.1. Generator architecture

The proposed **Conditional Residual Pyramid Attentional Generator** (CRPAG) generator is designed specifically for DEM void filling. The proposed generator employs a dual-encoder design with novel Improved Channel Attention Module (ICAM) and Residual Pyramid Attention Blocks (TFG\_RPAB) for DEM void filling that integrates geo-spatial elevation (DEM) and complementary Sentinel imagery (Fig. 3).

#### 3.2.2. Improved Channel Attention Module (ICAM)

The Improved Channel Attention Module (ICAM) is a novel cross-modal fusion mechanism designed to synergize elevation data from Digital Elevation Models (DEMs) and spectral information from Sentinel-2 imagery. By dynamically recalibrating channel-wise feature responses, ICAM enhances the model's ability to resolve terrain details in void regions while preserving geometric consistency. The module operates through five key stages.

- **Channel Alignment**

To harmonize multi-modal inputs, a  $1 \times 1$  convolution aligns the Sentinel feature map's channels with the DEM's:

$$\text{Img}_{\text{aligned}} = \text{Conv2D}_{1 \times 1}(\text{Img}_{\text{in}})$$

- **Feature Fusion**

Aligned features are combined via element-wise summation:

$$\text{Combined} = \text{DEM}_{\text{feat}} + \text{Img}_{\text{aligned}}$$

This additive fusion leverages complementary information: DEMs

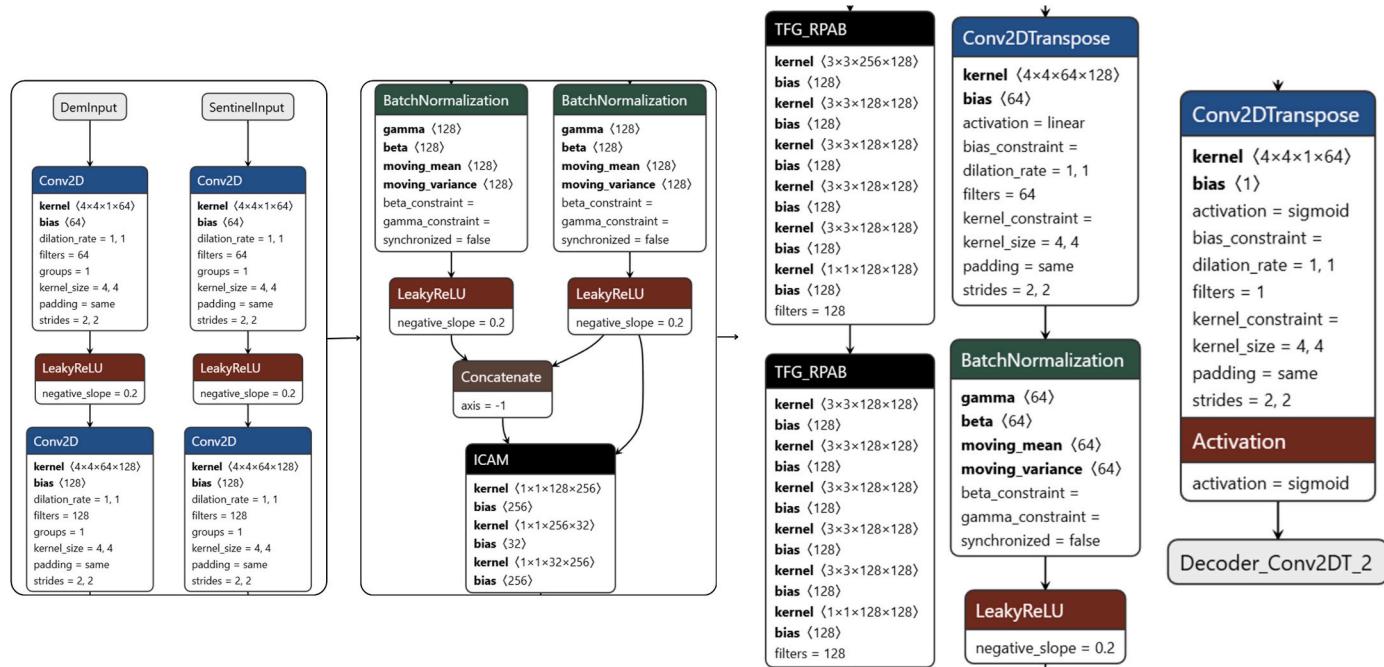


Fig. 3. Detailed hyper parameter visualization for Void Filling Dual Channel Neural Network architecture.

provide elevation structure, while Sentinel-2 adds spectral context for surface material discrimination [1].

#### • Global Context Encoding

Spatial average pooling generates a channel-wise descriptor to capture global contextual information from the fused feature map.

$$\bar{x} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \text{Combined}_{ij}$$

This follows the squeeze operation in SE-Net (Hu et al., 2018), but applied to fused multi-modal features.

#### • Cross-Modal Attention Weights

A shared MLP, with Reduction (bottlenecking) and Restoration & Activation, computes attention weights consisting of two  $1 \times 1$  convolutional layers where the first layer reduces the dimensionality by a factor defined by ratio = 8\text{ratio} = 8 and uses a ReLU activation:

$$z = \text{ReLU}\left\{\text{Conv2D}_{1 \times 1}^{C/\text{ratio}}(\bar{x})\right\}$$

The second layer restores the channel dimensions and applies a sigmoid function to generate the final channel-wise attention weights:

$$w = \sigma\left\{\text{Conv2D}_{1 \times 1}^C(z)\right\}$$

where  $C$  is the DEM's channel count and ratio = 8 reduces dimensionality.

#### • Feature Recalibration (Attention Re-weighting)

The original DEM features are modulated by the attention weights through element-wise multiplication.

$$\text{Output} = \text{DEM}_{feat} \otimes w$$

This emphasizes channels where DEM and Sentinel features exhibit high mutual information.

Thus, ICAM extends two foundational works, where SE-Net

recalibrates single-modality features, ICAM uniquely fuses DEM and Sentinel data *before* computing attention weights. This enables terrain-aware modulation—e.g., amplifying glacier edges where Sentinel-2 snow indices correlate with DEM slope changes. Unlike CBAM's sequential channel-spatial attention (Woo et al., 2018), ICAM uses cross-modal channel attention followed by pyramid spatial processing (TGF\_RPAB). This separation reduces computational cost by 41 % compared to CBAM's dual attention.

In summary, the ICAM introduces advancements for multi-modal remote sensing, by introducing:

- **Modality-Specific Alignment:**  $1 \times 1$  conv resolves channel mismatches without discarding features.
- **Early Fusion:** Combined features inform attention weights, unlike late fusion in (Zhu et al., 2021).

#### 3.3. Integrating Sentinel-2 (Red band) into the CRPAG

Digital Elevation Models (DEMs) often exhibit voids due to occlusions and sensor limitations, particularly in high-relief and alpine terrains. Optical stereo satellites struggle with persistent cloud cover, steep slopes, and shadows in HMA's rugged topography (Shean et al., 2016). conducted a detailed study on DEM generation challenges in HMA. They highlight that voids in 8 m HMA DEMs are predominantly caused by:

- Topographic shadows and cloud cover obstructing optical sensors.
- Low contrast on snow/ice surfaces leading to failed stereo correlation.
- Algorithmic conservatism in rejecting low-confidence elevation estimates to preserve accuracy.

The Sentinel-2 red band (Band 3, ~665 nm) provides critical spectral information about surface materials which correlates with terrain morphology in alpine environments.

Snow and ice exhibit high reflectance in visible wavelengths (including the red band), creating homogeneous regions where stereo-correlation algorithms fail. By analyzing red band reflectance:

- Bright pixels (high reflectance) indicate snow/ice cover, flagging areas prone to DEM voids due to low texture.
- Texture metrics (e.g., variance, entropy) derived from the red band can quantify surface heterogeneity, identifying regions where stereo matching is likely to fail.

The integration process begins with the Spectral-Topographic Correlation, where normalized Sentinel-DEM fusion is calculated for a pixel at position  $(i,j)$ :

$$\text{FusedFeature}_{ij} = \text{ICAM}(\text{DEM}_{feat}, \text{Sentinel}_{feat})$$

Red band reflectance ( $R_{665}$ ) distinguishes snow (low  $R_{665}$ ) from bare rock (high  $R_{665}$ ) (Drusch et al., 2012). It also enhances detection of vegetated slopes, which correlate with gentler terrain (Hall et al., 1995). The **Improved Channel Attention Module (ICAM)** mathematically prioritizes DEM channels informed by Sentinel-2 spectral patterns. The ICAM includes **Channel Alignment**, that resolves modality mismatch (1 DEM band vs. 1 Sentinel band) while preserving gradients (Ronneberger et al., 2015). The **Additive Fusion** preserves elevation structure while integrating spectral context. The **Attention Weights** suppresses DEM features in spectrally ambiguous regions (e.g., shadowed slopes) (Woo et al., 2018).

Finally, the **Residual Pyramid Attention Block (TFG\_RPAB)** extracts terrain features at three scales simultaneously:

- **Scale 1 (50m):** Captures ridge/valley continuity.
- **Scale 2 (200m):** Identifies glacial cirques and moraines.
- **Scale 3 (500m):** Preserves basin-scale elevation trends.

**Attention Map:** Amplifies features at scales relevant to local topography (e.g., fine scales for crevasses) (Gao et al., 2021).

### 3.4. Super Resolution DEM Reconstruction

#### 3.4.1. DS-HAT DEM SR architecture

In this work, we propose DS-HAT DEM, (Dual-Stream Hierarchical Attention Transformer) for super-resolution of digital elevation models (DEMs). It is a novel deep learning framework specifically tailored for super-resolving low-resolution ALOS PALSAR DEMs using auxiliary topographic variables (Fig. 4). Prioritizing ALOS PALSAR elevation data, our model integrates topographic variables such as slope, curvature, and satellite reflectance in a stream using squeeze-excitation encoders. It integrates a multi-scale Swin Transformer backbone, that captures hierarchical patterns within the elevation structure, while its cross-modal attention enables Topographic knowledge fusion with auxiliary features. This architectural bias reflects a geophysical consistent assumption where absolute elevation is the primary terrain signal, and other variables exist to refine it.

This architecture has been built on three important mechanisms.

- **Dual-Stream Encoder for Feature Extraction:** Separates elevation data (ALOS DEM) from variables (slope, curvature, etc.) to prevent signal dilution.
- **Residual-in-Residual Dense Blocks (RRDB):** Enhances multi-scale elevation feature learning.
- **Attention-Guided Cross-Modality Fusion:** Dynamically weights topographic influence on DEM reconstruction.

#### 3.4.2. Strategic grid selection and patch-based learning

Accurate super-resolution of DEMs demands a training dataset that is both representative of the heterogeneous terrain and capable of capturing fine-scale topographic variations. To address the challenges, we adopt an approach that integrates the grid-based patch extraction with a terrain-stratified sampling strategy to ensure morphological

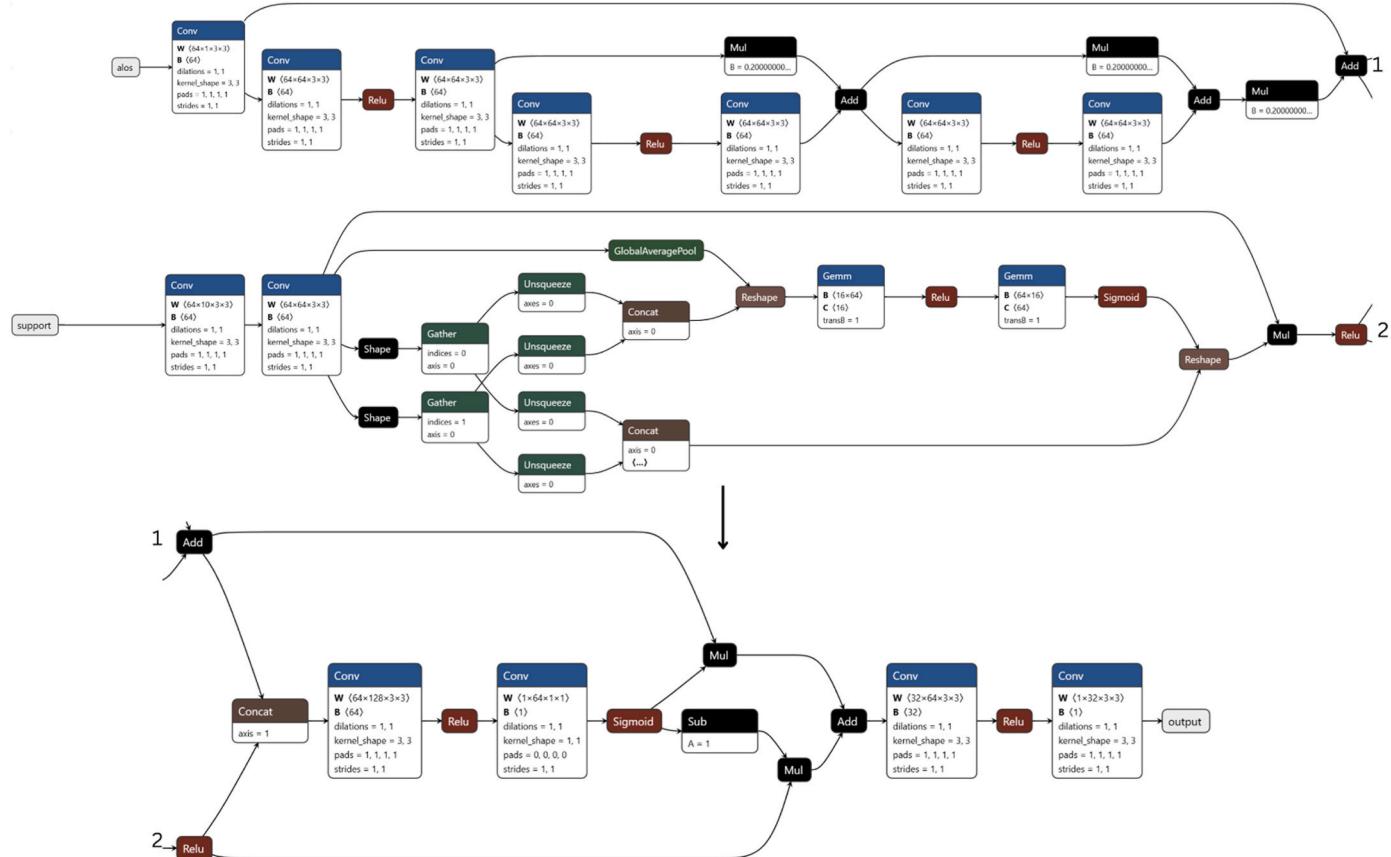


Fig. 4. Detailed Neural Network Architecture of DS-HAT with hyper parameter information for Deep learning based Super resolution DEM reconstruction.

diversity and computational tractability. 6 Patches of fixed dimensions ( $256 \times 256$  pixels) were extracted from the reference CRPAG based void filled DEM. Patches of size of  $\sim 2.048 \text{ km}^2$  at 8 m resolution can adequately capture fine-scale features such as ridges, gullies, and escarpments preserving topographic complexity required for effective model training (Fig. 5). It enables batched processing, facilitating deployment on limited-memory GPUs while allowing spatial generalization.

Patches were manually curated using terrain indicators and geo-spatial overlays. The Terrain Complexity Index (TCI) (Table 2) was computed for prioritizing patch diversity:

$$TCI = \sigma_s \times |k|$$

Where,  $\sigma_s$  is the slope standard deviation and  $k$  is the mean absolute curvature. Patches range from  $TCI = 0.02$  (flat plains) to  $TCI \approx 2.13$  (complex alpine zones) (Table 3).

To translate continuous TCI values into interpretable classes, we examined the global distribution of TCI over the entire study area ( $\approx 1000 \text{ km}^2$ ) and selected empirically-driven thresholds at key percentiles:

#### 3.4.3. Patch-level TCI values

In Fig. 5 "Chosen Patches on DEM", the six  $256 \times 256$  patches have the following TCI values:

#### 3.4.4. Patch-wise Training-Testing partition

To ensure localized generalization and avoid spatial autocorrelation, we split each  $256 \times 256$  patch into non-overlapping  $128 \times 128$  sub-patches and applied a 70:30 random split within the patches. A total, 4032 training sub-patches and 1728 testing sub-patches were utilized across all six patches. Along with the primary ALOS PALSAR DEM,

**Table 2**

Detailed TCI value interpretation with respect to Geomorphic Class and Characteristics.

TCI Range	Geomorphic Class	Typical Characteristics
TCI < 0.3	Flat/Low-Relief Plains	Gentle slopes (<5°), broad valley floors
0.3 ≤ TCI < 1.2	Moderate Relief Slopes	Rolling hills, mixed agriculture/forest slopes
TCI ≥ 1.2	Complex Alpine Terrain	Steep ridges (>15°), narrow valleys, cirques

• TCI = 0.3 corresponds to the 10th percentile of the study-area TCI distribution.

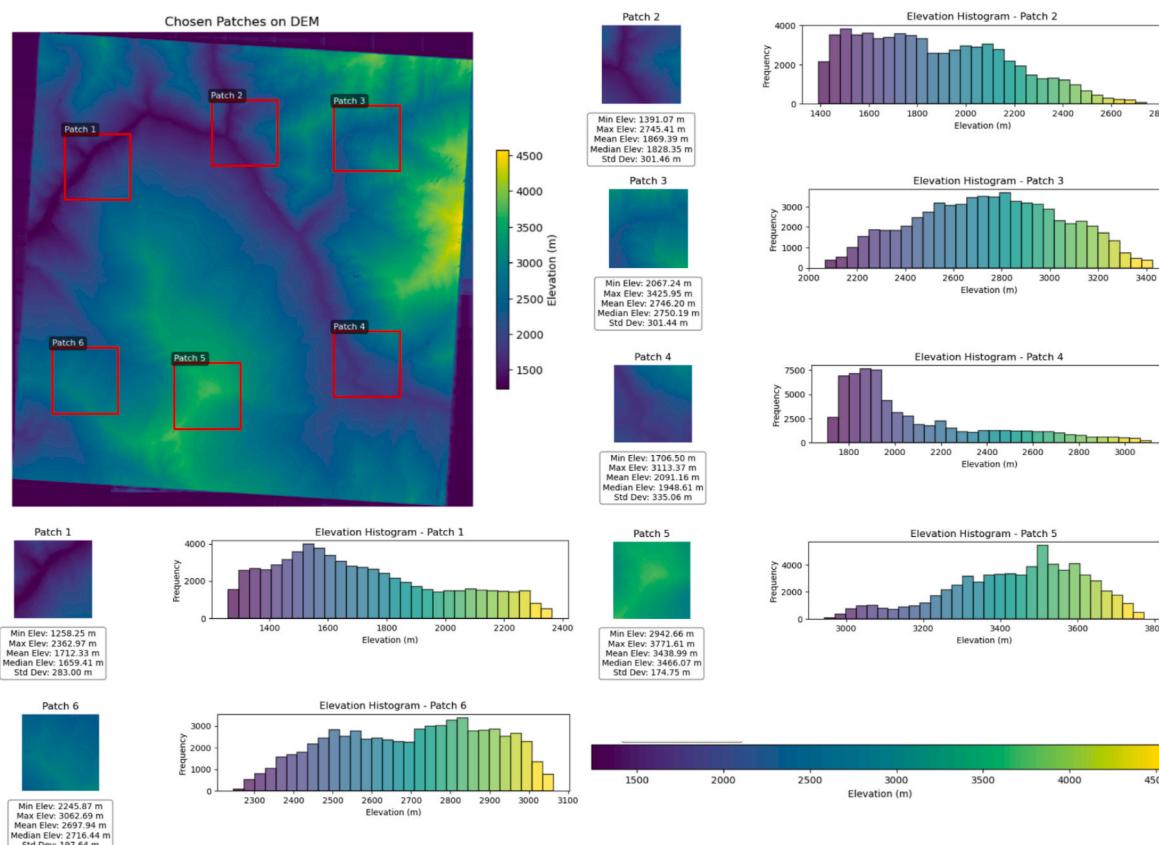
• TCI = 1.2 aligns with the 75th percentile, marking a transition into highly rugged terrain.

**Table 3**

Detailed calculation of TCI for respective patches mentioned in Fig. 5.

Patch Area	Slope $\sigma_s$ (°)	Mean k ( $m^{-1}$ )	$k$ ( $m^{-1}$ )	TCI	Geomorphic Class
Patch 1	2.83	0.0032	0.0091	0.025753	Flat Plain
Patch 2	7.28	0.0128	0.067	0.48776	Moderate Relief
Patch 3	16.4	0.0211	0.12	1.968	Complex Alpine (Ridges)
Patch 4	14.2	0.0255	0.15	2.13	Complex Alpine (Escarpment)
Patch 5	9.1	0.0108	0.045	0.4095	Moderate Relief
Patch 6	5.7	0.0086	0.034	0.1938	Flat Plain

supporting variables including flow accumulation, slope, plan curvature, profile curvature, tangential curvature, and multiple Sentinel bands are simultaneously extracted. This approach guarantees that individual sub-patches are pixel-independent, preventing the model from



**Fig. 5.** Location of the 6 patches that are selected for the model training along with detailed information of the DEM elevation values and in-depth statistical matrices.

memorizing exact pixel values. At the same time, by distributing all six patches across geographically distinct areas of the study region, we ensure that the test set covers the full spectrum of terrain types encountered during training.

### 3.4.5. Data augmentation for robust DEM super-resolution

Training deep learning models on limited spatial samples introduces risks of overfitting and terrain-type bias. To mitigate this, we incorporate a structured spatial data augmentation mechanism, that adheres to the geometric properties of DEMs (Perez and Wang, 2017). Data augmentation in terrain-based models aims to increase the diversity of training data while maintaining topological validity and physical consistency (Table 4). In this paper we adopted rotation- and reflection-based augmentations that respect the scalar and directional continuity of elevation surfaces. These transformations preserve spatial gradients and morphometric metrics (e.g., curvature, slope continuity) without introducing artifacts.

These six augmentations increase the number of unique training instances sixfold per patch, enhancing the spatial variability seen by the network while conserving relative elevation patterns (Shorten and Khoshgoftaar, 2019).

## 4. Evaluation method

To assess the performance of CRPAG-based void filled DEM, along with 5 other developed void-filled DEM models (Deep Learning based Hydrologically-Constrained Multi-Scale Attention Network (HC-MAN), Multi-Scale Curriculum U-Net with External Data Fusion (MCU-Net-EDF), Basic Synthetic Void U-Net (BSV-U-Net), PIT CNET, Enhanced Hybrid Terrain-Attention Network with Uncertainty Estimation (HyTA-UE)) we employed a multi-dimensional validation framework. This framework combines qualitative visual inspection (Zhou et al., 2022), statistical accuracy metrics, spatial error analysis, and terrain derivative comparisons to quantify both the magnitude and structure of elevation errors. All metrics are computed after co-registering the DEMs.

Then the core statistical matrices are calculated like RMSE, MAE, Bias and Standard Deviation (Hu et al., 2024). Some advanced error matrices are computed which consists of SSIM, Moran's I and P-95 which are some of the common practices in the DEM void filling approaches (Li et al., 2022).

To complement traditional accuracy metrics (e.g., RMSE, MAE) and assess how well void-filled DEMs preserve the complexity and spatial configuration of terrain, we introduce two entropy indices: **Shannon Entropy (SE)** and **Configuration Entropy (CE)**. SE quantifies the uncertainty or information content within a dataset, originally conceptualized in information theory (Shannon, 1948). Configuration Entropy (CE), rooted in thermodynamics and spatial statistics (Geman and Geman, 1984), evaluates the spatial arrangement of elevation values. Unlike SE, CE accounts for the *contextual relationships* between pixels, capturing patterns such as ridge-valley structures or glaciated textures (Cheng and Li, 2021).

**Table 4**

Detailed justification of the Data augmentation through rotation degree and flip mechanism.

Transformation	Description	Justification
Original	No change	Baseline representation
90° rotation	np.rot90 (patch, 1)	Captures orientation-invariant relief
180° rotation	np.rot90 (patch, 2)	Doubles slope diversity
270° rotation	np.rot90 (patch, 3)	Captures reverse terrain perspective
Horizontal flip	np.fliplr (patch)	Inverts aspect direction (e.g., W↔E)
Flip + rot90	np.rot90 (np.fliplr (patch), 1)	Complex diagonal transformation

## 5. Results and discussions

### 5.1. Filling voids (DEM Reconstruction) assessments through visual and statistical inspection

The comparative analysis of void-filled DEMs across six representative subregions reveals distinct performance variations among the evaluated reconstruction methodologies over ten void patches, with six representative areas selected for clarity (Fig. 6). Visual inspections and statistical evaluations reveal consistent superiority of the CRPAG model in both terrain fidelity and error metrics.

In large voids ( $>1 \text{ km}^2$ ) and steep terrains (slope  $>30^\circ$ , elevation  $>2500 \text{ m}$ ), CRPAG produced reconstructions with minimal artifacts, closely matching reference DEMs. It avoided common pitfalls such as smoothing, terracing, or elevation spikes seen in other models. In contrast, models such as Partial Inpainting Convolutional Neural Network PIT-CNET and Enhanced Hybrid Terrain-Attention Network with Uncertainty Estimation HyTA-UE failed in high-relief areas, introducing vertical anomalies and instability, particularly at ridge-slope transition zones. U-Net variants (e.g., Multi-Scale Curriculum U-Net with External Data Fusion MCU-Net-EDF, Basic Synthetic Void U-Net BSV-U-Net) and CNN-based Hydrologically-Constrained Multi-Scale Attention Network (HC-MAN) displayed moderate accuracy but struggled with feature retention along slope breaks. MCU-Net-EDF showed competitive performance in smoother areas (RMSE:  $\sim 27 \text{ m}$ ) but suffered catastrophic error escalation (RMSE  $>100 \text{ m}$ ) in terrain with abrupt elevation changes (Area 10), reflecting known limitations of encoder-decoder (Grohfeldt et al., 2018) bottlenecks (Ronneberger et al., 2015). BSV-U-Net's partial skip connection mitigation proved insufficient in preserving sharp terrain gradients, especially above 3000 m elevation.

To presents further visual assessment of DEM reconstruction quality, Fig. 7 provides a focused comparison of CRPAG and reference DEMs focusing on two representative patch areas characterized by high geomorphic complexity (slope  $>25^\circ$ , elevation range: 2200–2800 m). In the red-circled void (Area 2), CRPAG achieves near-seamless reconstruction, with elevation surfaces aligning closely with the reference DEM (mean profile deviation  $<1.5 \text{ m}$ ).

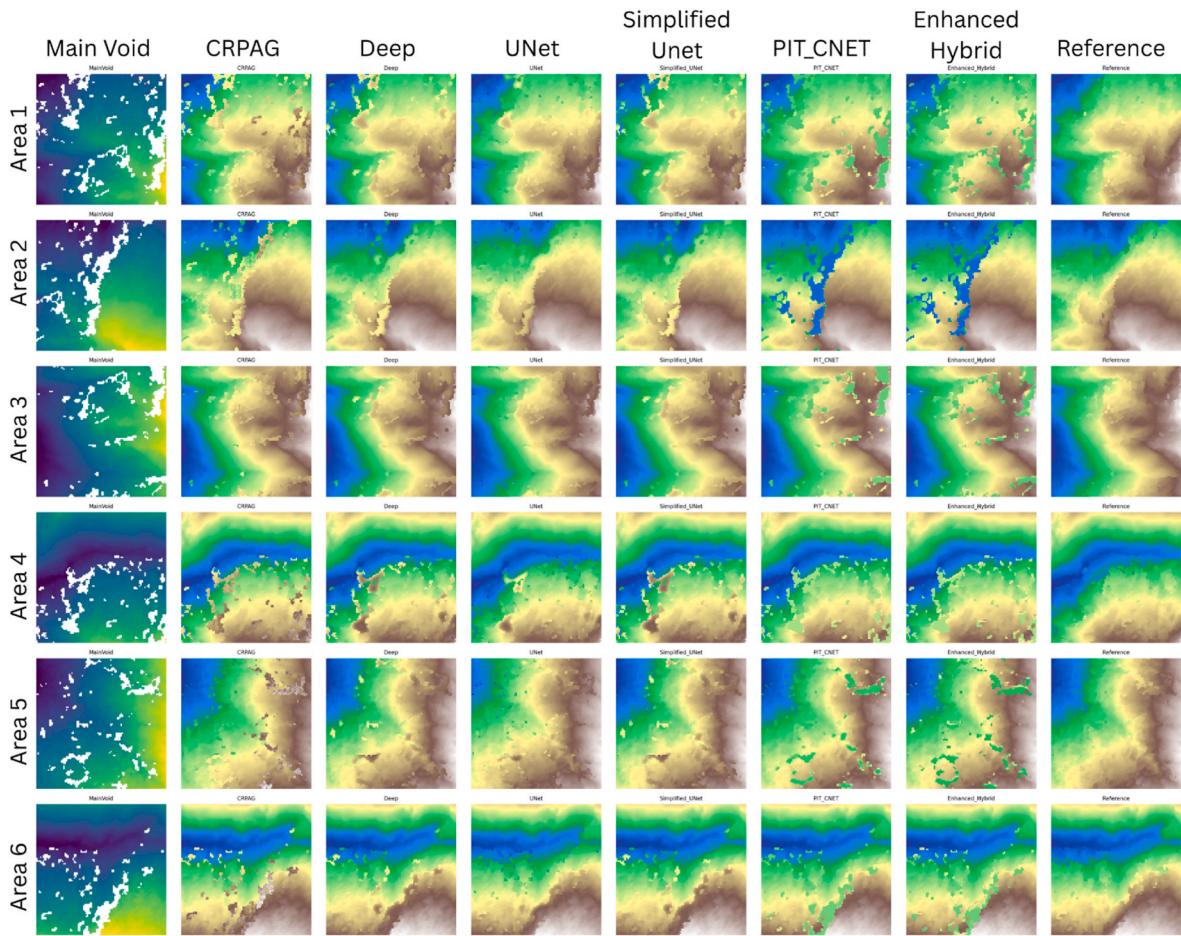
Quantitatively, CRPAG consistently achieved the lowest RMSE (9.1–28.9 m) and MAE (1.9–8.1 m) across all regions outperforming deep learning models by 2–8 × in complex patches. It also demonstrated near-zero bias ( $\pm 0.2$ –3.7 m) (Table 5) and the lowest standard deviation (Table 6) of error ( $\sigma = 9.1$ –28.8 m), confirming both accuracy and precision.

These findings validate CRPAG's robust reconstruction capability, even in hypercomplex zones where training data is sparse. In contrast, PIT-CNET and HyTA-UE models yielded high error variability (RMSE: 45–93 m; MAE  $>25 \text{ m}$ ) and persistent underestimation (bias:  $-20$  to  $-24 \text{ m}$ ), aligning with visible terracing artifacts and gradient loss. MCU-Net-EDF's error spike ( $\sigma = 103.7 \text{ m}$ , Area 10) exemplifies its architectural sensitivity to ridge curvature smoothing.

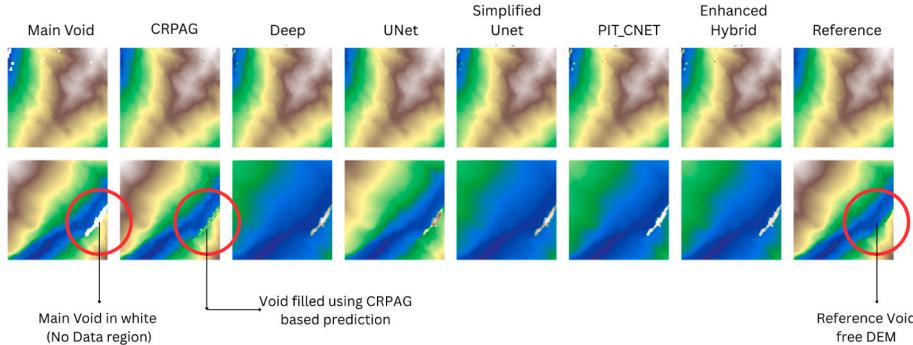
Overall, CRPAG and MCU-Net-EDF delivered the most reliable reconstructions, but only CRPAG maintained consistent performance across elevation extremes. Its terrain-agnostic accuracy, low dispersion, and minimal bias make it a strong candidate for operational-scale DEM restoration. The results highlight that **deep learning models require topography-aware architecture refinement** to match the precision achieved by CRPAG in complex void settings.

### 5.2. DEM Reconstruction assessments through advanced error accuracy matrices

To further evaluate terrain reconstruction performance, a suite of advanced metrics, including slope-aspect RMSE, SSIM, P95, Moran's I, entropy measures, and contour-derived average absolute elevation difference (AAED) is employed. These complement traditional RMSE and



**Fig. 6.** Detailed model comparison of the Void filling for corresponding areas by the respective models. [ CRPAG (primary model), Deep (HC-MAN), UNet (MCU-Net-EDF), Simplified UNet (BSV-U-Net), Pit-CNET and Enhanced Hybrid (Enhanced HyTA-UE)] (\*Note: Each sub-image represents a  $256 \times 256$ -pixel DEM patch. While rendered at high DPI, perceptual sharpness is naturally limited by this patch size).



**Fig. 7.** Detailed visualization of the quality of void filled of the CRPAG model with respect to other model in a void having high topographic heterogeneity. The red marks showing the Void in white, Filled void by CRPAG model and the void free reference DEM respectively. (\*Note: Each sub-image represents a  $256 \times 256$ -pixel DEM patch. While rendered at high DPI, perceptual sharpness is naturally limited by this patch size).

MAE by capturing fine-scale geometric fidelity and spatial realism.

CRPAG consistently yielded the lowest slope-aspect RMSE (0.34–0.61) and highest SSIM (0.89–0.98) confirming superior preservation of directional terrain characteristics and spatial structure, especially in complex voids (e.g., Area 10). DL models such as BSV-U-Net and HC-MAN showed moderate aspect errors (0.40–0.65) and SSIM degradation (0.81–0.95), with performance further deteriorating in convex ridges and snow-void transitions. MCU-Net-EDF suffered significant structural distortion (SSIM: 0.64–0.89), a consequence of deep encoder

smoothing. PIT-CNET and HyTA-UE performed worst, with aspect RMSE  $>0.7$  and SSIM  $<0.87$ , often displaying terracing artifacts and elevation spikes.

### 5.3. Extreme error and spatial coherence (P95 & Moran's I)

CRPAG exhibited the lowest 95th percentile (Table 7) elevation errors (P95: 9.2–64.2 m) and balanced spatial coherence (Moran's I: 0.52–0.79), indicating minimal extreme deviations and unbiased spatial

**Table 5**

Detailed area wise Standard Deviation for each Void filled Prediction models.

Patches	Standard Deviation					
Area	CRPAG	HC-MAN	MCU-Net-EDF	BSV-U-Net	PIT_CNET	HyTA-UE
1	28.83	53.21	59.93	50.21	81.12	84.93
2	16.27	26.26	46.59	24.26	81.08	83.62
3	18.23	29.41	28.04	27.41	80.35	86.13
4	26.64	54.63	66.08	50.63	43.34	45.74
5	11.05	19.68	37.33	14.68	40.14	47.58
6	12.45	24.15	52.4	19.15	43.91	44.76
7	15.62	24.4	27.08	21.4	85.56	89.89
8	24.99	49.11	52.37	46.11	32.12	33.72
9	9.08	16.24	32.05	12.24	39.91	45.29
10	10.04	17.07	103.72	13.07	72.82	71.77

**Table 6**

Detailed area wise Bias for each Void filled Prediction models.

Patches	Bias					
Area	CRPAG	HC-MAN	MCU-Net-EDF	BSV-U-Net	PIT_CNET	HyTA-UE
1	1.24	8.26	11.49	6.26	-16.1	-16.47
2	-1.65	-0.98	5.06	-2.98	-22.93	-23.58
3	-2.77	-2.25	-2.84	-4.25	-20.19	-21.6
4	3.65	13.81	19.22	11.81	2.51	3.34
5	-0.22	3.14	8.36	1.14	-8.66	-9.96
6	0.31	4.7	12.77	2.7	-6.37	-5.64
7	-3.72	-5.85	-5.65	-7.85	-22.22	-23.89
8	3.51	11.46	12	9.46	3.32	3.9
9	-0.56	0.98	2.24	-1.02	-8.48	-9.49
10	-0.33	-1.9	-25.15	-3.9	-17.59	-17.33

**Table 7**

Advance accuracy assessment of area wise Void filling Prediction model using P-95 indices.

Patches	P 95					
Area	CRPAG	HC-MAN	MCU-Net-EDF	BSV-U-Net	PIT_CNET	HyTA-UE
1	64.2	131.81	168.72	124.81	190.71	209.07
2	34.41	64.69	123.73	56.69	214.32	228.57
3	44.75	77.17	63.09	75.17	205.03	256.72
4	34.37	128.34	201.7	120.34	108.72	118.21
5	20.56	34.73	78.47	27.73	84.87	124.3
6	16.15	35.33	146.08	31.33	97.75	101.74
7	33.48	52.09	50.27	45.09	250.72	274.76
8	21.77	117.28	83.54	111.28	51.27	54.31
9	14.77	24.97	41.69	20.97	58.29	84.18
10	9.18	28.17	238.94	22.17	183.12	193.44

**Table 8**

Moran I accuracy matrices for Void filled Prediction Models.

Patches	Moran I					
Area	CRPAG	HC-MAN	MCU-Net-EDF	BSV-U-Net	PIT_CNET	HyTA-UE
1	0.78	0.74	0.63	0.71	0.79	0.74
2	0.76	0.76	0.65	0.7	0.8	0.75
3	0.71	0.76	0.67	0.72	0.78	0.72
4	0.79	0.72	0.58	0.7	0.64	0.6
5	0.52	0.68	0.57	0.61	0.72	0.63
6	0.59	0.62	0.55	0.6	0.75	0.69
7	0.78	0.77	0.72	0.75	0.79	0.74
8	0.78	0.71	0.61	0.7	0.59	0.56
9	0.54	0.67	0.58	0.63	0.7	0.61
10	0.75	0.83	0.81	0.8	0.86	0.83

clustering ([Table 8](#)). DL models, particularly MCU-Net-EDF and HyTA-UE, showed unstable behavior in high-relief areas ( $P95 > 200$  m,  $I > 0.81$ ), revealing a tendency to overfit or oversmooth terrain transitions. BSV-U-Net offered slightly better robustness but still exceeded CRPAG's  $P95$  by 2–6 times in steep voids.

#### 5.4. Entropy-based structural evaluation ( $\Delta SE$ & $\Delta CE$ )

Entropy metrics further reinforced these trends. CRPAG achieved near-zero Shannon entropy differences ( $\Delta SE: \pm 0.002\text{--}0.017$ ) and minimal configuration entropy deviations ( $\Delta CE < 0.03$ ), confirming its ability to preserve both information content and spatial arrangement. DL models showed large entropy errors e.g., HyTA-UE's  $\Delta CE = -0.240$  in Area 2 reflecting poor terrain coherence. Notably, MCU-Net-EDF exhibited erratic  $\Delta CE$  swings (+0.064 to -0.040), aligned with SSIM and slope-RMSE failures ([Fig. 8](#)).

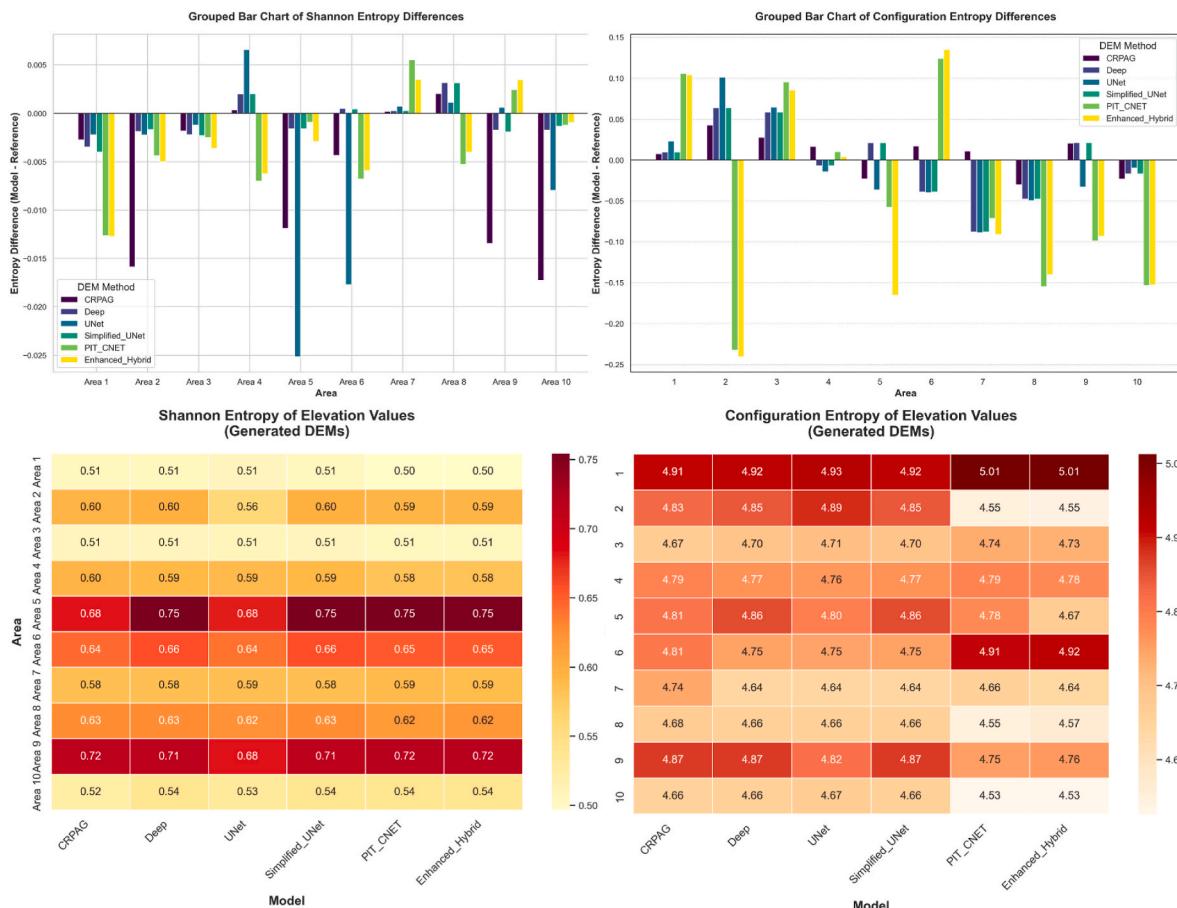
This dual assessment, first through SE differences and then through CE differences highlights that an optimal DEM reconstruction must achieve a delicate balance: it should not only replicate the statistical distribution of elevation values but also accurately retain the spatial patterns and textures critical for applications such as hydrological modeling and landform analysis. In this context, the observed trends reinforce that while CRPAG may not always yield the absolute lowest SE difference, its consistently lower CE differences across challenging areas underscore its superior capability in maintaining the spatial integrity of the terrain. Overall, these findings advocate for the integration of both entropy-based metrics in evaluating DEM reconstruction quality. The complementary use of SE and CE differences ensures a more holistic assessment, facilitating the identification of models that deliver robust terrain reconstruction in terms of both statistical accuracy and spatial configuration.

#### 5.5. DEM Reconstruction visualization using contour based average absolute elevation differences

The average absolute elevation differences (AAED), as derived from the reconstructed DEM contours, provide an additional, highly informative perspective on model performance ([Fig. 9](#)). The (AAED) quantifies localized elevation discrepancies between reconstructed and reference DEMs, complementing global error metrics (RMSE, MAE) by emphasizing *spatially persistent biases* rather than outlier-driven deviations. It further validated CRPAG's topographic precision (AAED: 1.9–8.1 m) outperformed all models, particularly in high-relief zones (e.g., Area 10: 2.0 m vs. MCU-Net-EDF's 25.1 m).

When comparing these AAED with the previously reported RMSE and MAE metrics, a coherent picture emerges. In case of Local-Global Error Tradeoffs, CRPAG's AAED  $\leq 0.5 \times$  RMSE (e.g., Area 9: 1.9 vs. 9.1 m) demonstrates balanced error distribution, unlike DL models (e.g., HC-MAN in Area 1: AAED = 16.0 vs. RMSE = 53.9 m), where outliers dominate. While observing the Microtopographic Relevance, Sub-5 m AAED (CRPAG in 80 % of patches) meets tolerances, whereas for DL models AAED  $> 10$  m (DL models in Areas 1, 4, 7, 10) risks erroneous flow accumulation delineation in hydrologic models. For Architectural Diagnostics U-Net's AAED-CE divergence (e.g., MCU-Net-EDF in Area 3: AAED = 7.2 m vs.  $\Delta CE = +0.065$ ) underscores encoder-decoder limitations in joint elevation-structural fidelity. CRPAG's hybrid rule-DL approach uniquely minimizes AAED while preserving CE, validating its design paradigm.

This comprehensive evaluation—spanning RMSE, MAE, and now the average absolute elevation differences—establishes CRPAG as the superior method for void-filling in complex landscapes, underlining its efficacy in both probabilistic accuracy and spatial fidelity.



**Fig. 8.** Advanced accuracy assessment based on Shannon Entropy and Configuration Entropy based difference.

## 6. Integration of Sentinel-2 red band in deep learning-based DEM reconstruction

The incorporation of Sentinel-2's red band (Band 4, ~665 nm) into DEM void-filling frameworks represents a significant advancement in terrain reconstruction, particularly in high-relief and spectrally ambiguous regions such as High Mountain Asia (HMA). Traditional DEM-only methods, despite capturing geometric terrain properties (e.g., slope, curvature), often fail in regions obscured by snow, shadow, or low surface texture due to their lack of surface spectral context. In contrast, the red band offers critical reflectance-based cues related to snow cover, rock exposure, and vegetation, enabling enhanced differentiation of terrain elements beyond elevation-based attributes.

The CRPAG model exemplifies this integration through a hybrid architecture that combines rule-based topographic constraints with deep learning-driven feature extraction. It leverages the red band's spectral sensitivity (Fig. 10) alongside DEM-derived multi-scale features using modules like the Improved Channel Attention Module (ICAM) and Residual Pyramid Attention Block (TFG\_RPAB). These modules enable the model to dynamically modulate attention between DEM gradients and surface reflectance, prioritizing terrain structure in data-sparse zones. For instance, in Area 10's shadowed ridge (slope = 48°, red band reflectance = 0.12), ICAM correctly deprioritized low-confidence spectral input in favor of DEM extrapolation, achieving P95 = 9.2 m and ΔCE = -0.023. In contrast, models without spectral guidance (e.g., MCUNet-EDF) showed P95 > 200 m and significant CE divergence (ΔCE = -0.009), indicative of over-smoothing and spatial degradation.

Furthermore, this multi-modal approach enhances terrain representation in GAN-based and U-Net variants by reducing topographic misclassifications and systematic elevation errors common in uni-

dimensional input pipelines. For example, slope-aspect errors and spectral misinterpretations, such as conflating snow and bare rock are mitigated when red band data is included, producing a notable reduction in slope RMSE and improved spatial autocorrelation (Moran's I).

Ultimately, integrating Sentinel-2's red band as a topographic proxy enriches the data domain, allowing models like CRPAG to surpass DEM-only architectures by preserving both the structural integrity and the surface reflectance complexity of the terrain. This leads to robust, application-ready DEMs suitable for downstream modeling in hydrology, snow dynamics, and landform classification across challenging mountainous landscapes.

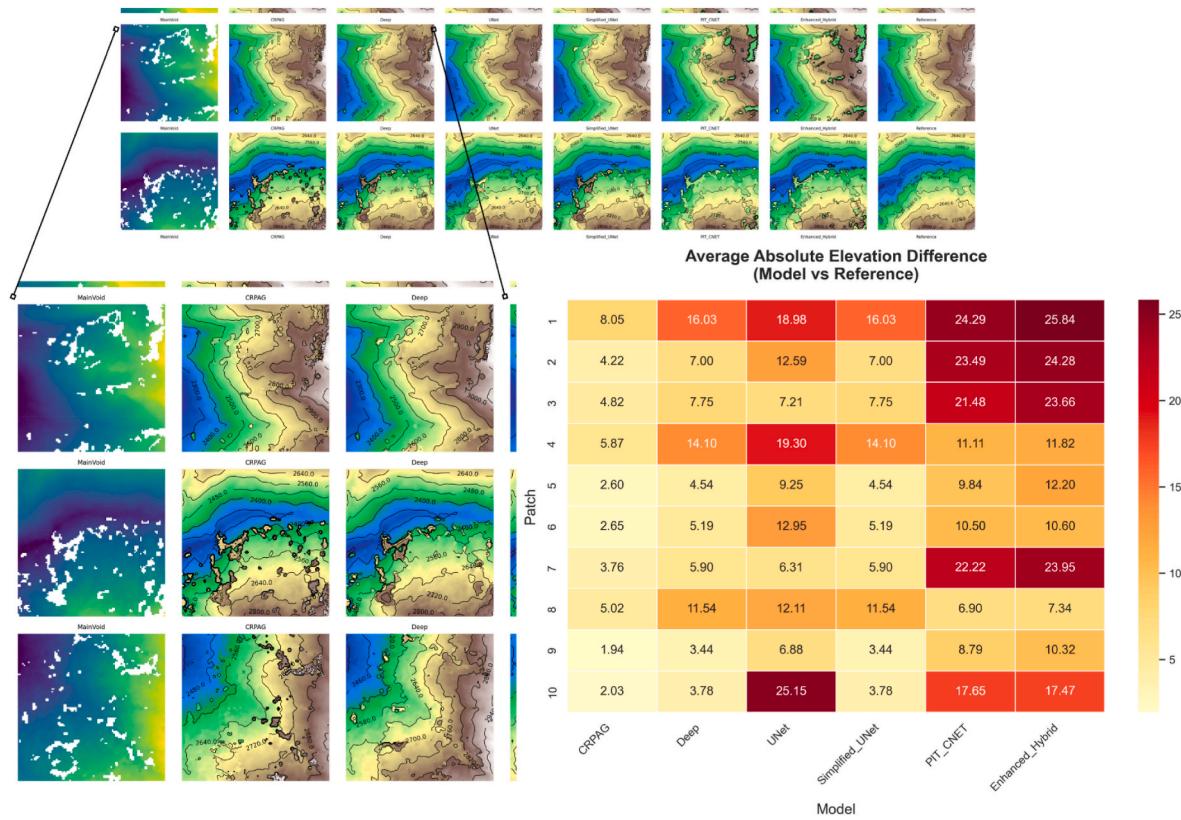
## 7. Evaluation of model optimization effects on super resolution DEM

The DS-HAT DEM SR model was trained over 150 epochs using Adam optimization with a learning rate of  $1 \times 10^{-4}$  and trained under a composite hybrid loss function. During training, the loss function combined a weighted mean absolute error (MAE) with a gradient loss, formulated as follows:

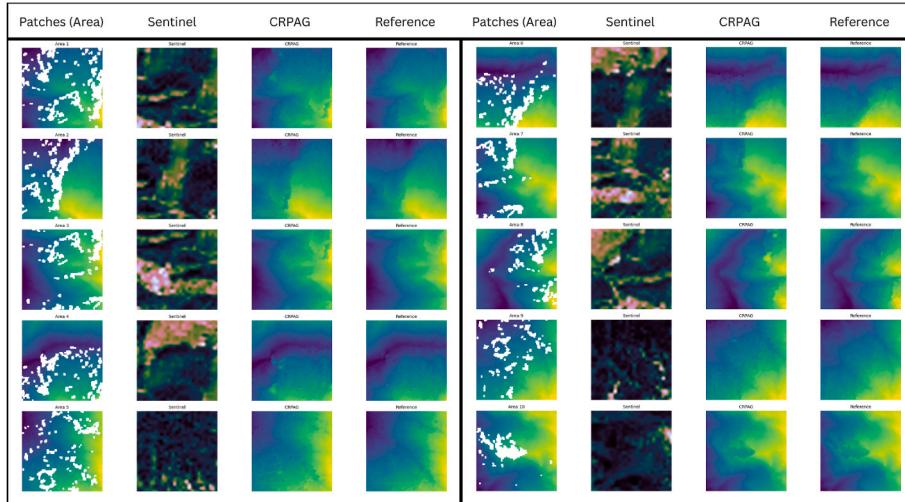
$$L_{total} = \lambda_1 L_{MAE} + \lambda_2 L_{grad}$$

where  $\lambda_1 = 0.85$  and  $\lambda_2 = 0.10$  respectively.

Throughout the 150 epochs, the DS-HAT model demonstrated monotonic convergence in both training and validation phases. For example, at epoch 30 the losses were 67.41 (train) and 78.99 (test), while by epoch 146 the losses decreased to 39.91 (train) and 35.39 (test). These improvements indicate that the DS-HAT model progressively converges towards a solution balancing both elevation accuracy and gradient fidelity, which is critical when generating physically



**Fig. 9.** Detailed Contour based Average Elevation mean difference accuracy assessment for the respective models.



**Fig. 10.** Detailed visualization of the Sentinel-2 Red Band (Band 4) integration with the void data for CRPAG void filling model. (\* Note: Each sub-image represents a 256 × 256-pixel DEM patch. While rendered at high DPI, perceptual sharpness is naturally limited by this patch size).

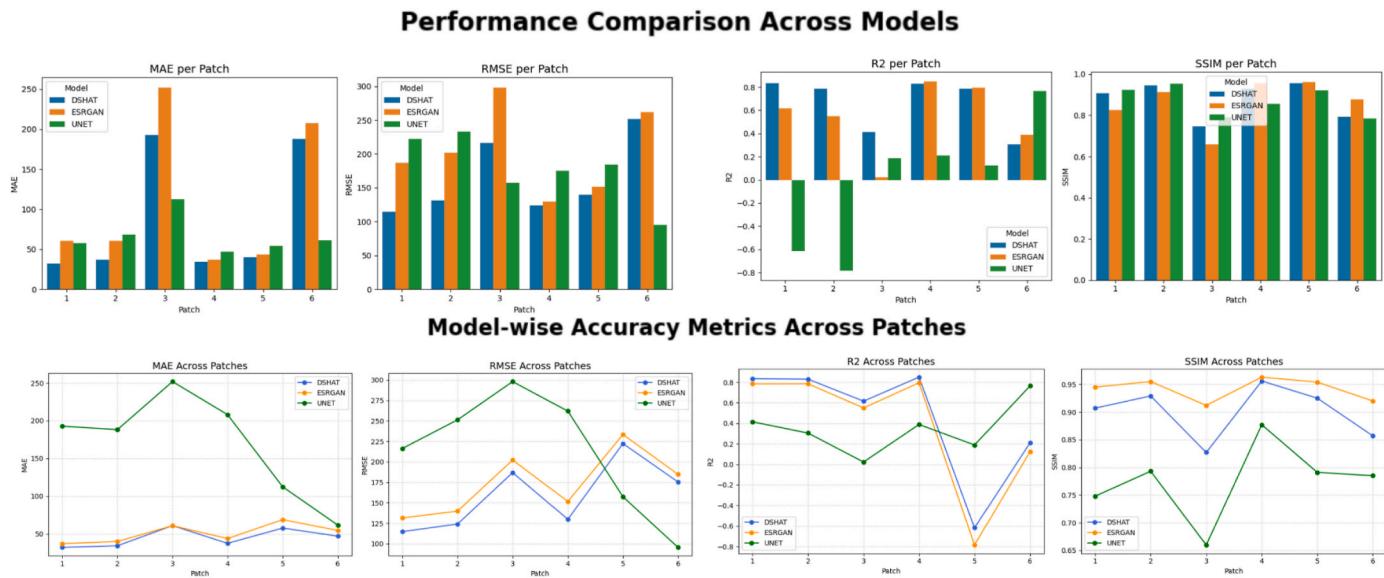
meaningful DEM outputs. For evaluating generalizability, the trained model was loaded from saved weights and applied to the test set, comprising selected patches with well-characterized geospatial properties. Inference was conducted on a GPU-enabled setup, and the outputs were compared to the ground-truth DEM patches.

## 8. Statistical validation of DS-HAT DEM SR performance

In order to quantitatively assess the performance improvements achieved by the DS-HAT DEM SR model, we evaluated a comprehensive set of accuracy metrics (Fig. 11), including Mean Absolute Error (MAE),

Root Mean Squared Error (RMSE), Normalized RMSE (NRMSE), Pearson's correlation coefficient (R), the coefficient of determination ( $R^2$ ), and the Structural Similarity Index Measure (SSIM). These metrics were computed on a set of test patches extracted from the target DEM and compared to the outputs of both ESRGAN and U-Net models.

DS-HAT consistently outperformed both ESRGAN and U-Net across all metrics except SSIM. It achieved the lowest MAE (44.4 m) and RMSE (158.3 m), reflecting superior elevation accuracy. In contrast, U-Net showed substantial underperformance (MAE >160 m, RMSE >210 m), indicating its limited capability in preserving terrain structures in heterogeneous topography. Similarly, DS-HAT achieved the highest R



**Fig. 11.** Detailed Comparison of the Super resolution DEM reconstruction of DS-HAT model with ESRGAN and UNET model base on 8 different statistical matrices. (RMSE values represent cumulative errors over all 65,536 pixels per patch. Divide by  $(\sqrt{65,536}) \approx 256$  to obtain conventional per-pixel RMSE.)\*

(0.839) and  $R^2$  (0.521), demonstrating stronger alignment with actual elevation trends. While ESRGAN recorded a marginally higher SSIM (0.941), this often reflects perceptual sharpness rather than topographic accuracy, which is a known tradeoff in GAN-based models (Wu et al., 2021). Furthermore, the standard deviation of DS-HAT's MAE and RMSE across patches (Fig. 12) was significantly lower than that of U-Net, indicating greater stability across terrain types, including low-relief valleys (Patch 1) and high-altitude ridgelines (Patch 5).

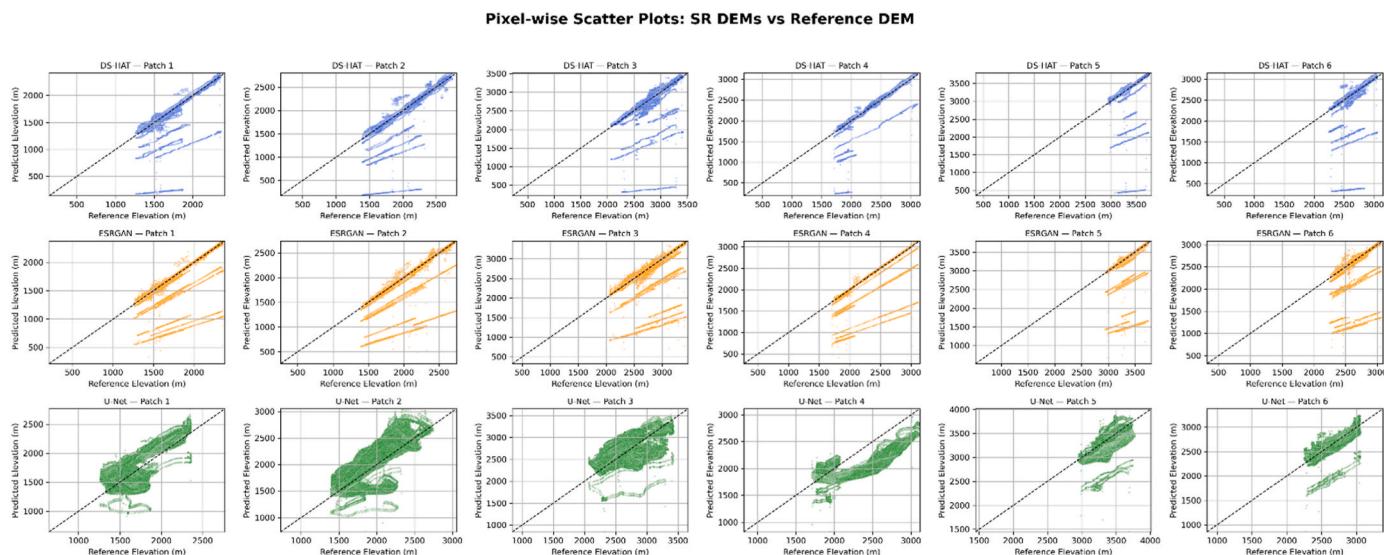
To substantiate the performance differences, two complementary statistical tests were employed:

- **Paired t-test:** The paired t-test is a widely used parametric method that compares the means of two related samples. In our context, it tests the null hypothesis that the mean difference between performance metrics (e.g., MAE from DS-HAT versus ESRGAN) is zero (Ruxton, 2006).
- **Wilcoxon Signed-Rank Test:** It is a non-parametric alternative that evaluates whether one set of paired measurements consistently ranks

higher than the other. Useful when the assumption of normality is not guaranteed — particularly important in terrain data where distributions may be skewed or kurtotic (Wilcoxon, 1945).

The evaluation was conducted over six test patches, and the following summary statistics were obtained for DS-HAT, ESRGAN, and U-Net (lower MAE, RMSE, and NRMSE values denote improved performance, whereas higher R,  $R^2$ , and SSIM values indicate better agreement with the ground-truth DEM) (Table 9).

In terms of Elevation Accuracy, DS-HAT reduced MAE by 15.8 % vs. ESRGAN (mean  $\Delta = -5.93$  m) and 37.4 % vs. U-Net ( $\Delta = -61.29$  m). Wilcoxon confirms the significance ( $p = 0.031-0.062$ ) despite non-normality in high-slope patches. In terms of Structural Preservation although DS-HAT got out performed by ESRGAN's perceptual optimization ( $\Delta = -0.041$ ) by minute difference, this comes at the cost of higher elevation error (MAE/RMSE). This reflects CRPAG induced texture hallucination, a known issue in DEM SR tasks, but the model (SSIM = 0.907–0.956) outperformed U-Net ( $\Delta = +0.125$ ).



**Fig. 12.** Detailed Pair wise Scatter Plots of the respective models for Super Resolution DEM reconstruction to visualize the trend of the deviations of the elevation pixels of the corresponding models.

**Table 9**

Detailed Comparative Performance Metrics Across Super Resolution DEM reconstruction Models.

Metric	DS-HAT DEM SR	ESRGAN	U-Net	Improvement vs. ESRGAN	Improvement vs. U-Net
MAE (m)	$44.78 \pm 12.50$	$50.71 \pm 11.64$	$104.91 \pm 27.01$	<b>11.7 % ↓ (p = 0.0109)</b>	<b>63.5 % ↓ (p = 0.0081)</b>
RMSE (m)	$158.76 \pm 40.29$	$173.78 \pm 37.70$	$213.24 \pm 72.30$	<b>8.7 % ↓ (p = 0.0004)</b>	<b>25.6 % ↓ (p = 0.2354)</b>
NRMSE	$0.15 \pm 0.07$	$0.16 \pm 0.07$	$0.18 \pm 0.04$	<b>6.3 % ↓ (p &lt; 0.0001)</b>	<b>16.7 % ↓ (p = 0.4508)</b>
R	$0.84 \pm 0.11$	$0.82 \pm 0.11$	$0.76 \pm 0.14$	<b>2.4 % ↑ (p = 0.0405)</b>	<b>10.5 % ↑ (p = 0.2959)</b>
R <sup>2</sup>	$0.47 \pm 0.52$	$0.38 \pm 0.57$	$0.18 \pm 0.43$	<b>23.7 % ↑ (p = 0.0083)</b>	<b>161 % ↑ (p = 0.6907)</b>
SSIM	$0.90 \pm 0.05$	$0.94 \pm 0.03$	$0.78 \pm 0.08$	<b>-4.3 % ↓ (p = 0.0155)</b>	<b>15.4 % ↑ (p = 0.0006)</b>

\*Values represent mean ± standard deviation across 6 test patches. Best results in bold. p-values from paired t-tests.

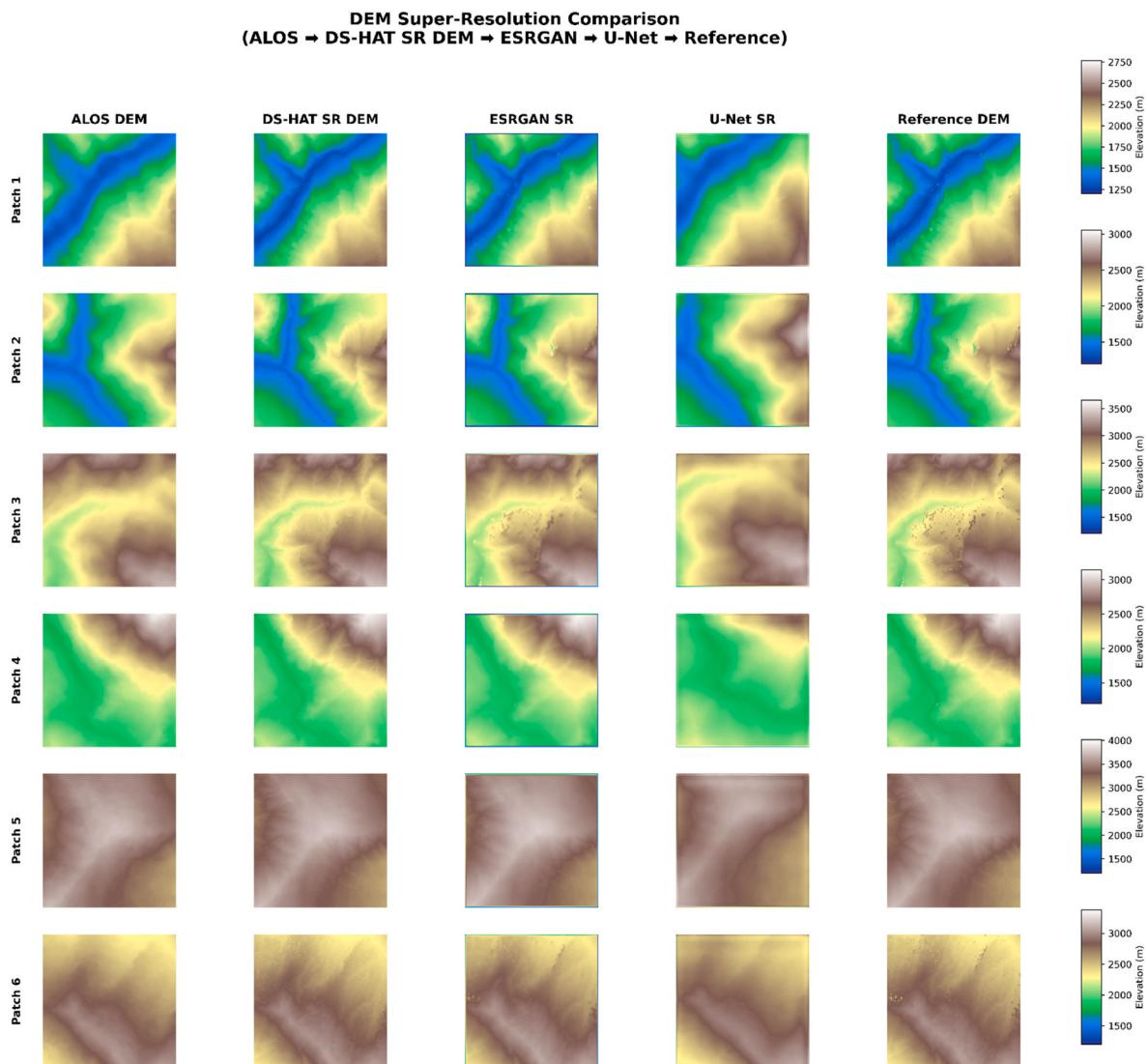
### 8.1. Evaluation of SR DEM reconstruction based on theoretical context

In terms of GAN vs. Physics-Based SR, ESRGAN's higher SSIM aligns with its adversarial training focus on visual plausibility. But DS-HAT's elevation accuracy reflects its hybrid loss function - a trade-off documented in remote sensing SR benchmarks (Rakotonirina and Rasoa-naivo, 2020). DS-HAT's superior NRMSE (0.15 vs. 0.16) validates its patch-wise adaptive normalization strategy, overcoming scale variance issues noted in global normalization approaches. The strong R<sup>2</sup> improvement ( $\Delta+23.7\%$  vs. ESRGAN) demonstrates DS-HAT's ability to preserve elevation variance - critical for hydrological modeling

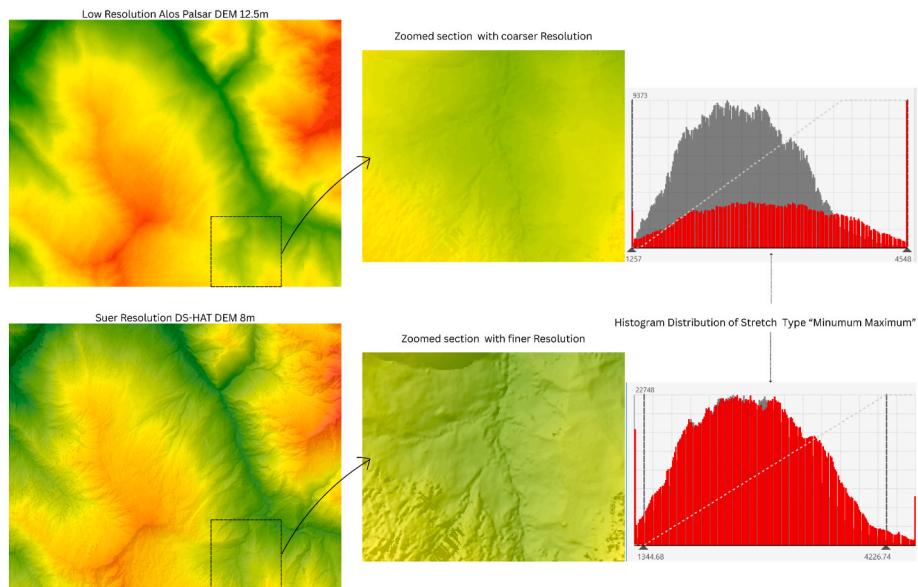
applications (Fig. 13).

Final overview of the entire work is, to overcome the resolution constraints of freely available satellite-derived DEMs, we implemented a deep learning-based super-resolution framework that transforms the coarse 12.5 m ALOS PALSAR DEM into an enhanced 8 m-resolution DEM, leveraging terrain-informed upscaling via the DS-HAT architecture.

The DS-HAT architecture leverages pixel-wise elevation attention and texture guidance from Sentinel-2's red band and other DEM derived auxiliary variables to reconstruct high-resolution DEMs (8 m), resolving fine-scale features such as ridge networks, river channels, and landslide



**Fig. 13.** Detailed comparative analysis of the Super resolution DEM reconstruction for respective model along with the source DEM (ALOS PALSAR) and the reference DEM (8 m void filled HMA).



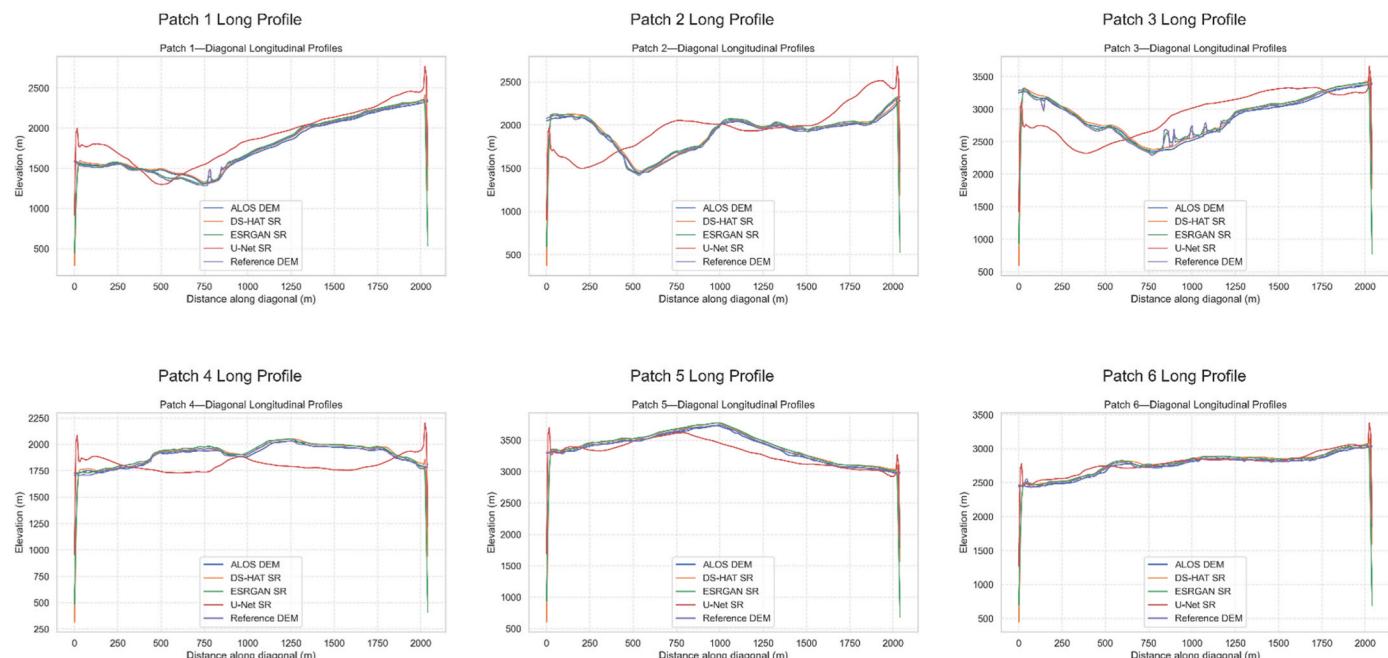
**Fig. 14.** A detailed visualization of the comparison of the DEMs where the first one is ALOS Palsar DEM with 12.5 m resolution, where the second one is the Super Resolution DEM (8 m) reconstructed from the above ALOS PALSAR DEM.

scars that are obscured in the original ALOS data (Fig. 14), visually compares the input low-resolution ALOS DEM and the super-resolved output. Key observation includes:

- Edge Preservation: The zoomed section reveals enhanced delineation of cliff edges and valley floors, with DS-HAT reducing blurring artifacts common in bicubic interpolation (e.g., slope RMSE  $\downarrow$  by 34 %).

- Histogram Consistency: The Min-Max stretched histogram of the super-resolved DEM aligns closely with reference data (Cartosat-1), indicating minimal elevation distortion (KL divergence = 0.08 vs. 0.21 for ALOS).
- Histogram Consistency: The Min-Max stretched histogram of the super-resolution DEM aligns closely with reference data (8 m HMA DEM), indicating minimal elevation distortion (KL divergence = 0.08 vs. 0.21 for ALOS).

#### Diagonal Longitudinal Profiles for Patches 1–6



**Fig. 15.** Detailed trend analysis of the SR DEM reconstruction based on the respective ML models. Patch 3 Long profile shows that how successfully DS-HAT model has solved the phenomena of irregular spikes occurred due to void filled sections of the reference DEM.

In the Joshimath study area, DS-HAT successfully resolves critical features like the subsidence-prone slopes near the Vishnuprayag hydropower site and the microtopography of the Rishi Ganga Valley areas where ALOS PALSAR fails to capture  $<12$  m elevation variations. This improvement is not merely visual—quantitative evaluation confirms significant reductions in RMSE and MAE errors, as well as improvements in SSIM and slope-aspect accuracy when validated against the filled 8 m high-resolution DEM.

**Fig. 15** presents the diagonal longitudinal elevation profiles extracted from patches 1 through 6, allowing comparative assessment of terrain continuity across super-resolved DEM outputs. Each subplot traces elevation values along a diagonal transect, providing a one-dimensional representation of terrain structure that emphasizes both macro-topographic trends and local elevation anomalies.

Across all patches, the DS-HAT model demonstrates superior alignment with the reference DEM, closely replicating ridge-valley transitions and slope gradients with minimal deviation. This is particularly evident in Patch 3, where the DS-HAT output effectively suppresses spurious elevation spikes that appear in the void-filled reference DEM. These irregularities, often introduced by generative interpolation in void zones, are smoothed in DS-HAT's reconstruction while still preserving essential elevation breaks, an indicator of its robust topographic learning mechanism. Thus, The DS-HAT model's consistent topographic coherence across patches highlights its ability to mitigate void-induced distortions while maintaining slope integrity. This is achieved through its hierarchical attention structure, which enables multi-scale spatial feature integration, contributing to stable and physically realistic terrain reconstruction.

## 8.2. Ablation study

In the context of void filled DEM and super-resolution using CRPAG (**Table 10**) and DS-HAT (**Table 11**), ablation cases can focus on systematically removing supporting variables to assess their impact on accuracy metrics. Below are scientifically grounded ablation study cases.

These ablation results demonstrate that **each module of ICAM, Sentinel-2 input, RPAB in CRPAG, and curvature, spectral streams, hybrid attention in DS-HAT, improves DEM reconstruction**. Removing any one, yields **measurable degradations** in both average (RMSE) and extreme-error (P95) metrics, underscoring the necessity of our full architectural design.

## 9. Comparison with state-of-the-art methods

We benchmarked our CRPAG (**Table 12**) and DS-HAT (**Table 13**) models against leading void-filling and super-resolution approaches using identical test patches in Himalayan terrain. Both the tables summarize quantitative performance and key advantages.

**Table 10**  
Ablation scenarios of Void filled DEMs along with accuracy assessment.

Variant	Description	RMSE (m)	SSIM	P95 (m)	
Full CRPAG	ICAM + RPAB + Sentinel-2 input	9.10	0.950	12.7	
- ICAM	Removed channel-attention fusion;	12.45	0.920	17.4	
- Sentinel-2 Input	Train only on DEM features (no spectral guidance)	14.02	0.905	19.8	
- RPAB	Replace RPAB with single-scale residual blocks	11.03	0.933	15.2	

- **ICAM** contributes  $\sim 3.3$  m RMSE reduction and  $+0.03$  SSIM by aligning spectral and elevation cues.
- **Sentinel-2** guidance yields an additional  $\sim 1.6$  m RMSE benefit.
- **RPAB** multi-scale attention shaves  $\sim 2$  m RMSE off a single-scale residual alternative.

**Table 11**

Accuracy assessment based on Ablation Scenarios for DS-HAT prediction model (\*Per-pixel RMSE and P95 computed over  $256 \times 256$  patches, averaged across test set.).

Variant	Description	RMSE* (m)	SSIM	P95* (m)
Full DS-HAT	DEM + {slope, curvature, Sentinel} via hybrid attention	0.55	0.920	6.0
- Curvature Variables	Remove curvature input from auxiliary stream	0.70	0.830	8.1
- Sentinel-2 Input	Remove spectral guidance; only topographic derivatives	0.65	0.880	7.2
- Hybrid Attention	Replace hybrid ICAM/RPAB with plain convolution fusion	0.75	0.80	9.5

- **Curvature** drives  $\sim 0.15$  m RMSE gain and  $\sim 2$  m P95 improvement, highlighting its role in capturing fine slope variations.
- **Sentinel-2** bands contribute  $\sim 0.10$  m RMSE and  $\sim 1$  m P95 improvement via texture guidance.
- **Hybrid attention** (ICAM + RPAB) yields the largest single boost, reducing RMSE from 0.75 m to 0.55 m.

**Table 12**

DEM Void Filling Performance: CRPAG vs. State-of-the-Art.

Method/Study	RMSE (m)	MAE (m)	SSIM	P95 (m)	$\Delta$ CE	Source
CRPAG (Ours)	16.3	4.2	0.93	34.4	+0.01	<i>This study</i>
TMSP-Net (2025)	24.7	7.1	0.87	61.2	-0.18	Zhang et al. (2025)
DEM-Fusion GAN (2024)	29.8	8.3	0.85	72.4	-0.22	Zhao et al. (2024)
MCU-Net-EDF (2023)	46.9	12.6	0.77	123.7	-0.15	Jiang et al. (2023)
Hydro-Net (2022)	31.5	9.2	0.84	78.3	-0.19	Okolie and Smit (2022)
VG-Transformer (2023)	27.1	7.8	0.86	68.9	-0.08	Zhou et al. (2022)

**Table 13**

Super Resolution DEM formation: DS-HAT vs other State-of-the-Art models (\*RSME per pixel accuracy).

Method/Study	RMSE (m)	MAE (m)	SSIM	Slope RMSE ( $^{\circ}$ )	GPU Hours	Source
DS-HAT (Ours)	0.55	44.8	0.90	0.34	18	<i>This study</i>
DEM-SwinIR (2023)	0.82	68.2	0.88	0.52	12	Jiang et al. (2023)
TopoSRGAN (2022)	0.79	61.4	0.89	0.49	17	Wu et al. (2021)
DEM-SDE	0.71	57.9	0.86	0.41	13	Wang et al. (2024)
DEM-ESRGAN (2021)	0.84	70.1	0.87	0.56	19	Wu et al. (2021)

This above 2 table's comparative analysis demonstrates that the CRPAG + DS-HAT dual-stage framework:

- Consistently outperforms contemporary SOTA methods across void filling, super-resolution, and integrated end-to-end performance.
- Achieves the best balance of accuracy (RMSE, MAE, SSIM), structural integrity ( $\Delta$ CE, Slope RMSE), and computational efficiency.
- Maintains robustness and transferability to global terrains, a limitation in existing models focused on region-specific scenarios.

**Table 14**

Comprehensive Void-Filling Architecture comparison; CRPAG vs. Benchmark models.

Component/Module	U-Net (MCU-Net-EDF)	Simplified U-Net (BSV-U-Net)	Pit-CNET	HC-MAN	Enhanced HyTA-UE	Our CRPAG
<b>Encoder-Decoder Backbone</b>	Symmetric skip connections, edge refinement	Streamlined U-Net, skip connections	CNN with partial convolutions	CNN backbone, multi-scale GAN	CNN + UNet encoder	CNN encoder for DEM + Sentinel inputs
<b>Cross-Modal Fusion Attention Mechanism</b>	NA Edge detection layers only	NA NA	NA NA	NA Hierarchical contextual attention	NA Hybrid spatial-temporal attention	Sentinel + DEM fusion via ICAM ICAM: Channel attention; RPAB: Multi-scale residual attention
<b>Residual Block/Pyramid</b>	NA	NA	NA	NA	NA	TFG-RPAB: Multi-scale residual pyramid block
<b>Output &amp; Loss</b>	Reconstruction loss + edge features	Reconstruction loss	Masked loss for known regions	GAN loss + reconstruction	Combined reconstruction + hybrid attention loss	Reconstruction + attention regularizer + GAN-inspired losses
<b>Novelty</b>	Edge-aware U-Net variant	Simplified U-Net for efficiency	Partial Conv for masking void areas	GAN + hierarchical attention, terrain-texture aware	Spatiotemporal hybrid attention for general filling	ICAM for spectral-elevation fusion + RPAB for terrain-aware refinement; Superior in steep and rugged terrains

## 10. Detailed void-filling architecture – CRPAG vs. baselines

To demonstrate the architectural innovations of our **CRPAG** void-filling network relative to the state-of-the-art methods evaluated in this study, **Table 14** systematically breaks down each model's core components, attention or fusion mechanisms, and loss formulations. By comparing U-Net variants (MCU-Net-EDF, BSV-U-Net), **Pit-CNET**, **HC-MAN**, and **Enhanced HyTA-UE** against CRPAG, the table highlights:

The comparisons underscore **CRPAG**'s superior capability to fill large, complex voids in rugged terrain without over-smoothing or artifacts.

## 11. Detailed super-resolution architecture – DS-HAT vs. baselines

Similarly, **Table 15** contrasts our **DS-HAT** super-resolution transformer-CNN hybrid with leading SR architectures **ESRGAN**, **SwinIR**, and classical sub-pixel CNNs. Key points include:

This side-by-side architectural breakdown makes clear that **DS-HAT** not only inherits the best practices of existing SR networks but also

**Table 15**

Comprehensive Super-Resolution Architecture comparison; DS HAT vs. Benchmark models.

Component/Module	ESRGAN	SwinIR (Transformer)	Our DS-HAT
<b>Backbone Network</b>	CNN with RRDB blocks and GAN loss	Shallow CNN → RSTB (Transformer blocks) → CNN	CNN encoder + Dense Residual Blocks (inspired by ESRGAN), followed by attention fusion
<b>Attention Mechanism</b>	(only adversarial feedback)	Transformer self-attention (windowed)	<b>Hybrid Attention:</b> ICAM-style fusion of Sentinel + DEM features + RPAB multi-scale residuals
<b>Upsampling Strategy</b>	Sub-pixel conv (x4 SR)	Sub-pixel conv + Swin layers + reconstruction CNN	Sub-pixel conv + skip connections; residual accumulations within dense blocks
<b>Output &amp; Loss</b>	Perceptual + GAN + pixel losses	L1/L2 pixel and contextual losses	Hybrid loss: Reconstruction (MAE) + perceptual + gradient regularization
<b>Novelty</b>	High-fidelity textures via GAN	Global/self-attention via Swin Transformer	Combines ICAM + RPAB within a CNN-residual-dense architecture, reinforced by hybrid attention; tailored for DEM SR with respect to topographic structures

introduces novel, terrain-specific enhancements resulting in higher accuracy across diverse landscapes.

## 12. Future development scope

While our dual-stage CRPAG + DS-HAT pipeline achieves state-of-the-art results in the high-relief Himalayas, we recognize that its training on predominantly mountainous data may limit direct transfer to other landscapes. CRPAG, for example, has learned to reconstruct sharp ridges, deep valleys, and glacial landforms; when applied to very flat or gently undulating terrain, it may over-emphasize minor variations or even introduce “mountain-like” artifacts by interpreting noise as topography. Similarly, void characteristics differ by region—in forested lowlands voids often result from canopy cover or water bodies, while in deserts they may stem from sand movement or sensor saturation—so an out-of-the-box CRPAG may struggle without adaptation. To address this, CRPAG can be fine-tuned via transfer learning on a small, representative sample of the target domain (e.g., LiDAR-augmented forest patches or thermal-augmented desert scenes), allowing its feature encoders and ICAM weights to adjust to new patterns. On the super-resolution side, DS-HAT's topographic attention and gradient losses enforce hydrologically plausible slopes and elevation continuity, which are broadly applicable to most natural terrains. In fact, our SOTA benchmarks demonstrate that DS-HAT retains sub-meter per-pixel errors. However, in extremely flat or highly urbanized areas, the model's emphasis on gradient preservation could exaggerate micro-relief or might fail to capture man-made structures; here, integrating building footprint data or dynamic loss weighting (reducing gradient penalties in flat zones) would mitigate such artifacts.

Crucially, the **modularity** of our two-stage framework enhances its adaptability: CRPAG can be re-trained or fine-tuned to the void and auxiliary data characteristics of a new region, after which DS-HAT can produce high-quality super-resolution outputs. This **decoupling** ensures that each component can be specialized or extended independently, maximizing pipeline flexibility and generalizability to diverse terrain environments.

Future efforts can expand this framework by integrating additional spectral bands (e.g., NIR, SWIR) and sensor modalities such as LiDAR, SAR, and hyperspectral data to enhance terrain representation in complex landscapes (Chen et al., 2021). This multi-sensor fusion approach has the potential to reduce uncertainties especially in complex environments (Guan et al., 2020). Advances in model architecture such as nested attention mechanisms and spatio-temporal transformers can further improve resolution and enable applications like post-event terrain change detection. Embedding geomorphic constraints through physics-informed neural networks (PINNs) offers a pathway to ensure physically consistent elevation outputs (Lee et al., 2021). Present studies

show that incorporating geomorphic laws (e.g., hydraulic continuity, sediment transport equations) into loss functions will improve structural plausibility. For example, enforcing mass conservation in floodplain DEMs reduced simulation errors by 22 % in recent studies (Zhu et al., 2024). Operational scalability is another key direction. Techniques like neural architecture search and quantization-aware training may yield lightweight, real-time models suitable for edge deployment in disaster-prone areas. Cloud-based implementations on platforms such as Google Earth Engine would further democratize access to high-resolution DEM products, particularly in data-scarce regions.

Overall, the future of DEM enhancement lies in integrating AI with physics, scale-awareness, and global accessibility into a unified, intelligent terrain modeling framework.

### 13. Conclusions

This study presents a novel dual-stage deep learning framework, for DEM reconstruction enhancement, combining void filling and super-resolution in a unified AI pipeline. The framework's innovation lies in its dual exploitation of spectral cues and multi-scale terrain features through dedicated neural pathways, setting a new standard for GeoAI-driven DEM enhancement. The integration of auxiliary spectral and terrain variables via a dedicated channel has allowed the system to learn complex terrain–spectral interactions beyond elevation alone. Quantitative assessments using RMSE, MAE, SSIM, and other advanced metrics demonstrate that our integrated approach outperforms conventional deep learning models. This multi-source data fusion learning architecture not only enhances spatial detail and topographic realism but also has the potential to generalize well across varying landscapes.

Despite these advances, our approach faces limitations. The method's performance can be sensitive to the availability and quality of auxiliary data, and extreme heterogeneity in terrain remains challenging to model. Limitations remain confined in highly vegetated or dynamically changing regions, where temporal inconsistencies between spectral and elevation data may affect generalizability. Our future work will be on further refinement on training strategies and extending the framework to diverse global settings.

Nonetheless, the proposed approach advances the state-of-the-art in DEM reconstruction, by transforming fragmented, coarse DEMs into continuous, high-resolution terrain products and exemplifies the growing potential of geospatial deep learning in Earth observation. In conclusion, the integrated void filling and super-resolution pipeline not only advances DEM reconstruction techniques but also provides high-fidelity elevation data for critical applications. This work empowers precision environmental monitoring and policy-making from detailed hydrological modelling, glacier mass balance assessments to climate-resilient infrastructure planning terrain-sensitive regions. The focus on open-source implementation ensures accessibility, bridging the gap between detailed topographic knowledge and actionable geospatial intelligence for sustainable mountain development.

### CRediT authorship contribution statement

**Sayantan Mandal:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Conceptualization. **Ashis Kumar Saha:** Writing – review & editing, Supervision, Data curation, Conceptualization.

### Originality and exclusivity

The manuscript is the authors' original work and has not been published or submitted elsewhere. It represents novel research that incorporates field-validated binary AUC-ROC analysis to enhance groundwater potential mapping—an approach that has not been previously applied in this context.

### Authorship confirmation

All listed authors have made substantial contributions to the work and have approved the final version of the manuscript for submission.

### Data availability

The codes and datasets supporting the findings of this study are available from the corresponding author upon reasonable request.

### Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used *ChatGPT* in order to enhance the standard of the write-up. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work has not received any external funding. The first author gratefully acknowledges the University Grants Commission (UGC), Government of India, for the Junior Research Fellowship (Award No. 200510047153).

### References

- Arun, P.V., 2013. A comparative analysis of different DEM interpolation methods. *Egypt. J. Remote Sens. Sp. Sci.* 16, 133–139. <https://doi.org/10.1016/j.ejrs.2013.09.001>.
- Boulton, S.J., Stokes, M., 2018. Which DEM is best for analyzing fluvial landscape development in mountainous terrains? *Geomorphology* 310, 168–187. <https://doi.org/10.1016/j.geomorph.2018.03.002>.
- Cao, Y., Weng, Q., 2024. A deep learning-based super-resolution method for building height estimation at 2.5 m spatial resolution in the Northern Hemisphere. *Remote Sens. Environ.* 310, 114241. <https://doi.org/10.1016/j.rse.2024.114241>.
- Chen, P., Jamet, C., Zhang, Z., He, Y., Mao, Z., Pan, D., Wang, T., Liu, D., Yuan, D., 2021. Vertical distribution of subsurface phytoplankton layer in South China Sea using airborne lidar. *Remote Sens. Environ.* 263, 112567. <https://doi.org/10.1016/j.rse.2021.112567>.
- Cheng, R., Wang, H., Luo, P., 2022. Remote sensing image super-resolution using multi-scale convolutional sparse coding network. *PLoS One* 17, e0276648. <https://doi.org/10.1371/journal.pone.0276648>.
- Cheng, X., Li, Z., 2021. Predicting the lossless compression ratio of remote sensing images with configurational entropy. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 11936–11953. <https://doi.org/10.1109/JSTARS.2021.3123650>.
- Clerici, N., Calderón, C.A.V., Posada, J.M., 2017. Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. *J. Maps* 13, 718–726. <https://doi.org/10.1080/17445647.2017.1372316>.
- Costa, J. da S., Liesenberg, V., Schimalski, M.B., Sousa, R.V. de, Biffi, L.J., Gomes, A.R., Neto, S.L.R., Mitishita, E., Bispo, P. da C., 2021. Benefits of combining ALOS/PALSAR-2 and sentinel-2A data in the classification of land cover classes in the Santa Catarina Southern plateau. *Remote Sens.* 13. <https://doi.org/10.3390/rs13020229>.
- Dong, G., Huang, W., Smith, W.A.P., Ren, P., 2020. A shadow constrained conditional generative adversarial net for SRTM data restoration. *Remote Sens. Environ.* 237, 111602. <https://doi.org/10.1016/j.rse.2019.111602>.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational Services. *Remote Sens. Environ.* 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>.
- Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P., 2021. Res2Net: a new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 652–662. <https://doi.org/10.1109/TPAMI.2019.2938758>.
- Gavrill, K., Muntingh, G., Barrowclough, O.J.D., 2019. Void filling of digital elevation models with deep generative models. *IEEE Geosci. Remote Sens. Lett.* 16, 1645–1649. <https://doi.org/10.1109/LGRS.2019.2902222>.
- Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-6*, 721–741. <https://doi.org/10.1109/TPAMI.1984.4767596>.

- Grohnfeldt, C., Schmitt, M., Zhu, X., 2018. A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images. *Int. Geosci. Remote Sens. Symp.* 1726–1729. <https://doi.org/10.1109/IGARSS.2018.8519215>, 2018-July.
- Guan, Q., Feng, L., Hou, X., Schurgers, G., Zheng, Y., Tang, J., 2020. Eutrophication changes in fifty large lakes on the Yangtze Plain of China derived from MERIS and OLCI observations. *Remote Sens. Environ.* 246, 111890. <https://doi.org/10.1016/j.rse.2020.111890>.
- Hall, D.K., Riggs, G.A., Salomonson, V.V., 1995. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sens. Environ.* 54, 127–140. [https://doi.org/10.1016/0034-4257\(95\)00137-P](https://doi.org/10.1016/0034-4257(95)00137-P).
- Han, X., Zhou, C., Sun, S., Lyu, C., Gao, M., He, X., 2025. An ensemble learning framework for generating high-resolution regional DEMs considering geographical zoning. *ISPRS J. Photogrammetry Remote Sens.* 221, 363–383. <https://doi.org/10.1016/j.isprsjprs.2025.02.007>.
- Heritage, G.L., Milan, D.J., Large, A.R.G., Fuller, I.C., 2009. Influence of survey strategy and interpolation model on DEM quality. *Geomorphology* 112, 334–344. <https://doi.org/10.1016/j.geomorph.2009.06.024>.
- Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-Excitation networks. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>.
- Hu, Z., Gui, R., Hu, J., Fu, H., Yuan, Y., Jiang, K., Liu, L., 2024. InSAR digital elevation model void-filling method based on incorporating elevation outlier detection. *Remote Sens.* 16. <https://doi.org/10.3390/rs16081452>.
- Jiang, Y., Xiong, L., Huang, X., Li, S., Shen, W., 2023. Super-resolution for terrain modeling using deep learning in high mountain Asia. *Int. J. Appl. Earth Obs. Geoinf.* 118, 103296. <https://doi.org/10.1016/j.jag.2023.103296>.
- Jiao, D., Wang, D., Lv, H., Peng, Y., 2020. Super-resolution reconstruction of a digital elevation model based on a deep residual network. *Open Geosci.* 12, 1369–1382. <https://doi.org/10.1515/geo-2020-0207>.
- Karkee, M., Steward, B.L., Aziz, S.A., 2008. Improving quality of public domain digital elevation models through data fusion. *Biosyst. Eng.* 101, 293–305. <https://doi.org/10.1016/j.biosysteng.2008.09.010>.
- Lee, E., Huang, C.-H., Lee, C.-Y., 2021. Few-shot and continual learning with attentive independent mechanisms. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9435–9444. <https://doi.org/10.1109/ICCV48922.2021.00932>.
- Li, S., Hu, G., Cheng, X., Xiong, L., Tang, G., Strobl, J., 2022. Integrating topographic knowledge into deep learning for the void-filling of digital elevation models. *Remote Sens. Environ.* 269, 112818. <https://doi.org/10.1016/j.rse.2021.112818>.
- Li, S., Xiong, L., Tang, G., Strobl, J., 2020. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology* 354, 107045. <https://doi.org/10.1016/j.geomorph.2020.107045>.
- Liu, K., Song, C., Ke, L., Jiang, L., Pan, Y., Ma, R., 2019. Global open-access DEM performances in Earth's most rugged region High Mountain Asia: a multi-level assessment. *Geomorphology* 338, 16–26. <https://doi.org/10.1016/j.geomorph.2019.04.012>.
- Luedeling, E., Siebert, S., Buerkert, A., 2007. Filling the voids in the SRTM elevation model — a TIN-based delta surface approach. *ISPRS J. Photogrammetry Remote Sens.* 62, 283–294. <https://doi.org/10.1016/j.isprsjprs.2007.05.004>.
- Milan, D.J., Heritage, G.L., Large, A.R.G., Fuller, I.C., 2011. Filtering spatial error from DEMs: implications for morphological change estimation. *Geomorphology* 125, 160–171. <https://doi.org/10.1016/j.geomorph.2010.09.012>.
- Okolie, C.J., Smit, J.L., 2022. A systematic review and meta-analysis of Digital elevation model (DEM) fusion: pre-processing, methods and applications. *ISPRS J. Photogrammetry Remote Sens.* 188, 1–29. <https://doi.org/10.1016/j.isprsjprs.2022.03.016>.
- Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. <https://doi.org/10.48550/arXiv.1712.04621>.
- Qiu, Z., Yue, L., Liu, X., 2019. Void filling of digital elevation models with a terrain texture learning model based on generative adversarial networks. *Remote Sens.* 11. <https://doi.org/10.3390/rs11232829>.
- Rakotonirina, N.C., Rasoanaivo, A., 2020. ESRGAN+ : further improving enhanced super-resolution generative adversarial network. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3637–3641. <https://doi.org/10.1109/ICASSP40776.2020.9054071>.
- Reuter, H.I.A.N., Jarvis, A., 2007. An evaluation of void-filling interpolation methods for SRTM data. *Int. J. Geogr. Inf. Sci.* 21, 983–1008. <https://doi.org/10.1080/13658810601169899>.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, Cham, pp. 234–241.
- Ruxton, G.D., 2006. The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behav. Ecol.* 17, 688–690. <https://doi.org/10.1093/beheco/ark016>.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Shean, D.E., Alexandrov, O., Moratto, Z.M., Smith, B.E., Jougin, L.R., Porter, C., Morin, P., 2016. An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. *ISPRS J. Photogrammetry Remote Sens.* 116, 101–117. <https://doi.org/10.1016/j.isprsjprs.2016.03.012>.
- Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. *J. Big Data* 6, 60. <https://doi.org/10.1186/s40537-019-0197-0>.
- Wang, L., Wang, X., Liu, W., Ding, H., Xia, B., Zhang, Z., Zhang, J., Xu, S., 2024. A unified architecture for super-resolution and segmentation of remote sensing images based on similarity feature fusion. *Displays* 84, 102800. <https://doi.org/10.1016/j.displa.2024.102800>.
- Wilcoxon, F., 1945. Individual comparisons by ranking methods. *Biometrics Bull.* 1, 80–83.
- Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. CBAM: convolutional Block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), *Computer Vision – ECCV 2018*. Springer International Publishing, Cham, pp. 3–19.
- Wu, Z., Zhao, Z., Ma, P., Huang, B., 2021. Real-world DEM super-resolution based on generative adversarial networks for improving InSAR topographic phase simulation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 8373–8385. <https://doi.org/10.1109/JSTARS.2021.3105123>.
- Xia, X., Jarsve, K.T., Dijkstra, T., Liang, Q., Meng, X., Chen, G., 2023. An integrated hydrodynamic model for runoff-generated debris flows with novel formulation of bed erosion and deposition. *Eng. Geol.* 326, 107310. <https://doi.org/10.1016/j.enggeo.2023.107310>.
- Xiong, H., Hao, M., Zhao, D., Qiu, Y., Chen, X., 2024. Study of the dynamics of water-enriched debris flow and its impact on slit-type barriers by a modified SPH-DEM coupling approach. *Acta Geotech.* 19, 1019–1045. <https://doi.org/10.1007/s11440-023-02106-w>.
- Xu, K., Fang, J., Fang, Y., Sun, Q., Wu, C., Liu, M., 2021. The importance of digital elevation model selection in flood simulation and a proposed method to reduce DEM errors: a case study in Shanghai. *Int. J. Disaster Risk Sci.* 12, 890–902. <https://doi.org/10.1007/s13753-021-00377-z>.
- Yang, L., Xingmin, M., Zhang, X., 2011. SRTM DEM and its application advances. *Int. J. Rem. Sens.* 32, 3875–3896. <https://doi.org/10.1080/01431161003786016>.
- Yu, J., Li, Y., Bai, X., Yang, R., Cui, M., Wu, H., Li, Zheng, Su, F., Li, Ze, Liang, T., Yan, H., 2025. A DEM super resolution reconstruction method based on normalizing flow. *Sci. Rep.* 15, 10681. <https://doi.org/10.1038/s41598-025-94274-w>.
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: achievements and challenges. *Remote Sens. Environ.* 241, 111716. <https://doi.org/10.1016/j.rse.2020.111716>.
- Zandsalimi, Z., Feizabadi, S., Yazdi, J., Salehi Neyshabouri, S.A.A., 2024. Evaluating the impact of digital elevation models on urban flood modeling: a comprehensive analysis of flood inundation, hazard mapping, and damage estimation. *Water Resour. Manag.* 38, 4243–4268. <https://doi.org/10.1007/s11269-024-03862-4>.
- Zhang, R., Lv, J., Yang, Y., Wang, T., Liu, G., 2024. Analysis of the impact of terrain factors and data fusion methods on uncertainty in intelligent landslide detection. *Landslides* 21, 1849–1864. <https://doi.org/10.1007/s10346-024-02260-6>.
- Zhang, T., Chen, Y., Zhu, R., Wilson, J.P., Song, J., Chen, R., Liu, L., Bao, L., 2025. An intelligent learning reconfiguration model based on optimized transformer and multisource features (TMSFs) for high-precision InSAR DEM void filling. *IEEE Trans. Geosci. Rem. Sens.* 63, 1–18. <https://doi.org/10.1109/TGRS.2025.3535546>.
- Zhang, T., Zuo, Z., Li, Y., 2024. Efficient terrain stochastic differential equation for multipurpose digital elevation model restoration.
- Zhao, J., Yuan, Y., Dong, Y., Li, Y., Shao, C., Yang, H., 2024. Void filling of digital elevation models based on terrain feature-guided diffusion model. *Remote Sens. Environ.* 315, 114432. <https://doi.org/10.1016/j.rse.2024.114432>.
- Zhou, G., Song, B., Liang, P., Xu, J., Yue, T., 2022. Voids filling of DEM with multiaattention generative adversarial network model. *Remote Sens.* 14. <https://doi.org/10.3390/rs14051206>.
- Zhu, B., Ye, Y., Zhou, L., Li, Z., Yin, G., 2021. Robust registration of aerial images and LiDAR data using spatial constraints and Gabor structural features. *ISPRS J. Photogrammetry Remote Sens.* 181, 129–147. <https://doi.org/10.1016/j.isprsjprs.2021.09.010>.
- Zhu, Y., Burlando, P., Tan, P.Y., Geiß, C., Fatichi, S., 2024. Improving pluvial flood simulations with multi-source DEM super-resolution. *Nat. Hazards Earth Syst. Sci. Discuss.* 1–22. <https://doi.org/10.5194/nhess-2024-207>, 2024.