Terrain-Aware Attention Network for DEM Terrain Semantic Segmentation

Yuan Gao[1],Bo Zhou[1(✉)] and Yifan She[1]

Hefei University of Technology,Hefei 230009,China
yuan_gao@mail.hfut.edu.cn

Abstract Terrain segmentation is important for semantic extraction and scene understanding in geographic information systems. However, generic deep learning models lack mechanisms tailored to the physical characteristics of terrain data. We introduce the Terrain-Aware Attention (TAA) module, which uses terrain cues—local gradient, global elevation context, and surface ruggedness—to guide the attention mechanism for terrain-specific segmentation. We also propose the Hierarchical Asymmetric Attention Mechanism (HAAM), which assigns different attention modules to different network depths. We integrate TAA and HAAM into a U-Net architecture and evaluate them on a DEM dataset derived from ALOS PALSAR imagery. Experiments show that both TAA and HAAM independently improve segmentation performance over the baseline, and their combination achieves the highest IoU among all tested configurations.

Keywords: Semantic Segmentation; Attention Mechanism; Domain Knowledge; Digital Terrain

1 Introduction

Digital Elevation Models (DEMs) represent surface morphology and contain rich topographic information [16]. Unlike optical imagery, which is susceptible to atmospheric and lighting variations, DEMs reduce textural noise to reveal the underlying terrain structure. Thus, DEMs are widely used as base data for mountain segmentation.
Terrain semantic segmentation automatically assigns semantic labels to DEM pixels [23]. In this study, we consider a binary DEM segmentation task that separates mountain areas from surrounding lowlands for subsequent terrain analysis.

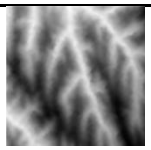
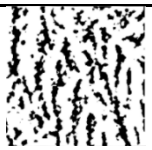
Visuals				
	DEM	Traditional	U-Net	**Ours**
Paradigm	Image	Thresh.	CNN	**TAA**
Connectivity	–	×	×	✓
Structure	–	×	Blurred	✓
Fine-grained	×	×	✓	✓
Dice Score	–	65.2 ↓**11.8**	77.0	**86.4** ↑**9.4**

Fig. 1. Comparison of terrain segmentation paradigms. Unlike traditional methods and standard CNNs, which suffer from fragmentation and discontinuities, our method preserves geometric structures. The Dice scores shown are computed on this single tile (randomly selected from the test set) and differ from the dataset-level averages reported in Table 2.

Traditional methods treat terrain data as standard grayscale imagery, using algorithms such as edge detection [22, 3, 8] and region growing [11] to extract topographic features for semi- or fully automated interpretation. These methods often produce fragmented results due to limited feature extraction. Deep learning models, in particular Convolutional Neural Networks (CNNs) and U-Net [14], have improved general segmentation, but applying them to terrain data remains challenging because few architectures account for DEM-specific topographic characteristics.

Elongated topographic structures such as ridgelines and valley floors are spatially continuous and highly distinctive. Standard U-Net architectures may break these structures during repeated down-sampling, creating discontinuities at boundaries such as canyon edges. Attention mechanisms are often added to improve features, but most are designed for generic feature recalibration and do not encode terrain descriptors or spatial autocorrelation patterns relevant to DEM interpretation. As a result, these models may miss physical terrain attributes and poorly delineate structures like ridgelines and valleys, as shown in Fig. 1.

Our problem differs from most attention work in one respect. Existing modules such as SE [7], CBAM [20], CA [6], and SimAM [24] are *domain-agnostic*: they recalibrate features from data statistics alone, without encoding terrain-related descriptors. Applied to DEM data, these modules treat elevation grids the same way they treat a natural photograph and do not model geophysical quantities such as slope, elevation, and surface ruggedness. Our approach encodes these terrain-specific priors directly in the attention computation, linking generic feature learning with domain-informed feature weighting.

We introduce the TAA module, inspired by the *First Law of Geography*[1]. TAA does not formally model spatial autocorrelation, but it implements the intuition that terrain context matters by computing attention maps from three geomorphological cues (slope,

[1] Tobler's First Law of Geography posits that everything is related to everything else, but near things are more related than distant things.

elevation, and ruggedness) and fusing them with adaptive weighting. We also propose HAAM, which assigns different attention mechanisms to different network depths based on their computational cost and the semantic scale of features at each level, instead of using the same module everywhere.

In summary, the contributions of this paper are as follows:

- We evaluate several attention mechanisms on DEM data and find that inserting a single generic module gives limited or inconsistent gains, while depth-specific assignment and terrain-aware design are more effective.
- We propose TAA, which computes attention weights from terrain priors (slope, elevation, and ruggedness) rather than from data-driven statistics alone. This design captures physical terrain attributes that standard modules do not represent.
- We introduce HAAM, which distributes different attention mechanisms across network depths based on efficiency and complexity, preserving local geometric detail in shallow layers while capturing global context in deep layers.

2 Related Work

2.1 Image Segmentation Based on U-Net

Traditional terrain segmentation methods rely on manual delineation and classification, which is labor-intensive and often inadequate for complex topography. Deep learning has enabled automated segmentation, with U-Net [14] becoming a common backbone for these tasks.

In recent years, improved U-Net variants have been widely used in geographic information science for semantic segmentation of remote sensing imagery [25]. Wang [18] integrated dual attention and multi-scale feature fusion to improve small object detection in marine remote sensing. Zhang [26], Jonnala [9], and Wu [21] show that hybrid architectures, optimized skip connections, and ecological index integration can improve accuracy and robustness in building extraction and water body segmentation.

However, these advances target optical remote sensing imagery, which has rich spectral and textural information. DEM data consists of a single elevation channel without the multi-dimensional features of optical data. The main challenge in terrain segmentation is to extract structural features from such constrained input.

2.2 Attention Mechanisms

Attention mechanisms allow models to prioritize relevant parts of the input and are widely used in Geographic Information Science (GIS) [5].

Different modules emphasize different aspects. Self-Attention (SA) [17] captures long-range dependencies but at high computational cost. SE [7] recalibrates channel-wise features via global pooling. CBAM [20] sequentially integrates spatial and channel dimensions, while CA [6] uses coordinate encoding to preserve directional structures.

SimAM [24] extracts local features through an energy-based objective without additional parameters, and Triplet Attention [13] enables cross-dimensional feature interactions.

Despite this diversity, all the above modules are *domain-agnostic*: their weights come from learned data statistics (e.g., global averages, variances, or pairwise similarities) rather than from physical descriptors. This generality works well for natural images but can be a limitation for DEM data, where terrain morphology is governed by well-defined geophysical quantities such as slope, elevation, and surface ruggedness that generic pooling or channel recalibration does not directly capture. Terrain-specific descriptors in the attention computation could produce more informative feature weighting for DEM segmentation, as explored in Section 3.

3 Method

3.1 Design of the Terrain-Aware Attention

The TAA module (Fig. 2) adds terrain cues to feature weighting. Let X denote the input feature map and Y the refined output; the three branch outputs are denoted M_{slope}, M_{elev}, and M_{rug}, respectively. The module uses three parallel branches to encode three complementary terrain attributes: local gradient, global elevation context, and surface ruggedness. We choose these three descriptors because they are common terrain variables in geomorphological analysis and each captures a different geometric aspect: slope measures local steepness, elevation encodes absolute altitude context, and ruggedness quantifies surface complexity. Alternative descriptors such as curvature and aspect are also informative but are higher-order derivatives of elevation that can be noisy at our data resolution and would add complexity; we leave them to future work. By decoupling and recombining features from these perspectives, TAA can separate geometric structures such as ridgelines and valleys without auxiliary supervision. Input feature maps are processed by the following three parallel branches.

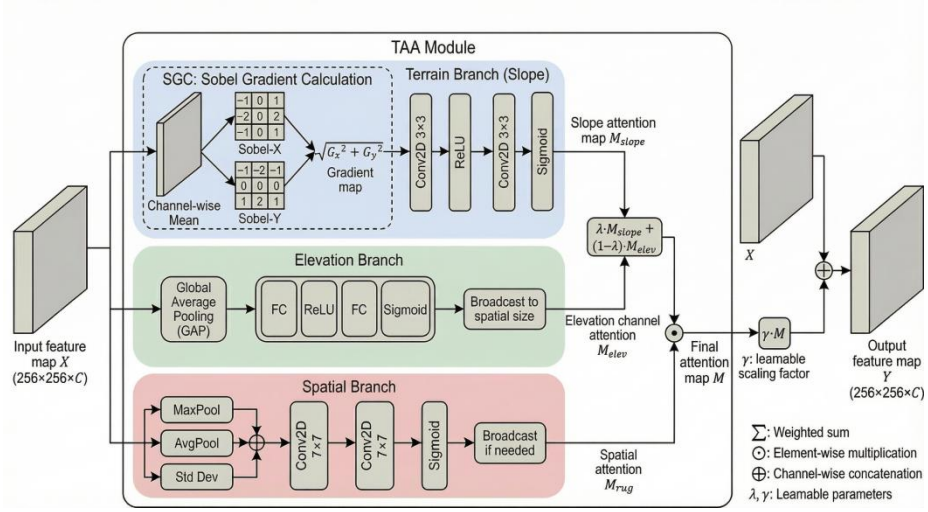


Fig. 2. The overall architecture of the TAA module. The input feature map is fed into three parallel branches. The Terrain Branch applies Sobel operators to generate gradient maps reflecting slope. The Elevation Branch uses Global Average Pooling to encode global altitude semantics. The Spatial Branch computes statistical descriptors and passes them through a large-kernel convolution to quantify surface ruggedness. The three outputs are fused via learnable parameters and to produce the enhanced feature map .

Terrain Branch. To capture slope-related local variation, we first compute the channel-wise mean of the input features. We then apply the Sobel operator to calculate the gradient magnitude in the horizontal G_x and vertical G_y directions. The slope attention map M_{slope} is generated through convolutional layers to highlight high-gradient areas:

$$G = \sqrt{G_x(\bar{X})^2 + G_y(\bar{X})^2}, \qquad (1)$$

$$M_{slope} = \sigma(\mathcal{F}_{conv}(G)), \qquad (2)$$

where $\bar{X}$ denotes the channel-mean feature, $\mathcal{F}_{\text{conv}}$ represents the stacked convolution layers with ReLU activation, and σ is the Sigmoid function.

Elevation Branch. Since different feature channels often respond to specific altitude layers, we employ Global Average Pooling (GAP) to encode global elevation context. A multi-layer perceptron (MLP) is used to generate channel-wise weights M_{elev}, emphasizing channels relevant to the global elevation distribution:

$$M_{elev} = \sigma(W_2 \cdot \delta(W_1 \cdot \text{GAP}(X))), \qquad (3)$$

where δ denotes the ReLU activation function.

Spatial Branch. To quantify surface ruggedness, we compute the maximum, mean, and standard deviation of the features along the channel dimension. These statistical descriptors are concatenated and processed by large-kernel convolutions to generate the ruggedness weight M_{rug}, guiding the network to focus on texturally complex regions:

$$X_{stat} = [\mathrm{Max}(X), \mathrm{Avg}(X), \mathrm{Std}(X)], \quad (4)$$

$$M_{rug} = \sigma(\mathrm{Conv}_{7\times7}(X_{stat})). \quad (5)$$

Adaptive Fusion & Output. We use a learnable coefficient λ to balance local geometric boundaries (slope) and global semantic context (elevation). The fused weights are then modulated by the ruggedness map to produce the final attention map M. The output Y is obtained via a residual connection with a learnable scaling factor γ:

$$\begin{aligned} M &= (\lambda \cdot M_{slope} + (1-\lambda) \cdot M_{elev}) \odot M_{rug}, \\ Y &= X + \gamma(X \odot M). \end{aligned} \quad (6)$$

Here, $\odot$ denotes element-wise multiplication. Both λ and γ are learnable scalar parameters updated via back-propagation, initialized to 0.5 and 0 respectively. Starting from $\lambda = 0.5$ assigns equal weight to slope and elevation, while $\gamma = 0$ lets the residual branch contribute nothing initially; the network then gradually learns the appropriate balance during training.

3.2 Attention Mechanism Combination Method

TAA improves the *content* of attention by adding terrain cues, but it does not address *where* different forms of attention should be placed in the network; these two aspects are complementary. TAA also has limited ability to extract multi-scale features. Prior work, including CoAtNet [4] and UniFormer [10], shows that representations differ by depth: shallow layers prioritize local textures and geometric details, while deep layers emphasize global semantics and context.

Based on this, we propose HAAM. Unlike methods that stack identical layers, HAAM selects an attention mechanism suited to each stage. We base our design on the efficiency and complexity analysis in Table 1. The resulting HAAM-UNet follows a standard encoder-decoder structure but places modules by stage.

Table 1. Efficiency and Complexity of Attention Mechanisms. "Param." denotes parameter increase; "Rec." indicates recommended stage.

Module	Param. Increase	Complexity	Rec. Stage
SA	≈ 0	$\mathcal{O}(n^2), n = H \times W$	Deep
SE	$2C^2/r$	$\mathcal{O}(C^2)$	Any
CBAM	$2C^2/r + 2k^2$	$\mathcal{O}(C^2 + H \times W)$	Any
CA	$2C^2/r + 2C$	$\mathcal{O}(C^2 + H + W)$	Middle
SimAM	0	$\mathcal{O}(C \times H \times W)$	Shallow
Triplet	$4C + 2k^2$	$\mathcal{O}(C \times H \times W)$	Deep

In the encoder, we match the module to the feature level. Shallow layers use the parameter-free SimAM to preserve textures without adding computational cost (see Table 1). Intermediate layers use CA to capture directional boundaries. Deep layers integrate CBAM and SE to encode complex semantics. For the bottleneck, we use a parallel

SE+CA structure to aggregate context. In contrast, the decoder uses an asymmetric design to reduce interference: CBAM is used only during deep recovery and SimAM at the final output, leaving intermediate layers attention-free.

3.3 Loss Function

The loss function measures the difference between model predictions and ground truth labels. U-Net architectures commonly use Cross-Entropy loss.
However, terrain data often has severe class imbalance between foreground and background; in our dataset the foreground-to-background ratio is approximately 3:7. Cross-Entropy loss tends to favor the majority class, causing the model to miss terrain details. Dice loss directly measures region overlap between predictions and ground truth, making it robust to class imbalance. We therefore adopt Dice loss.
Dice loss is derived from the Dice Similarity Coefficient [12], which measures the overlap between predicted and true segmentation maps on a scale from 0 to 1:

$$Dice\ coefficient = \frac{2|A \cap B|}{|A|+|B|}, \qquad (7)$$

where A is the prediction result, B is the true label, $|A \cap B|$ is the intersection area between the prediction and the true label, and $|A|$ and $|B|$ are the areas of the predicted region and the true label, respectively.
For optimization, we define the Dice loss as:

$$Dice\ Loss = 1 - Dice\ coefficient. \qquad (8)$$

4 Experiment

4.1 Dataset

We use the ALOS PALSAR DEM at 12.5 m spatial resolution. The data was acquired by the fully polarimetric PALSAR sensor aboard Japan's ALOS satellite via the NASA ASF platform. The study area is in a mountainous region of southern China, covering about 2,500 km^2 with elevations from tens of meters to over one thousand meters, including steep ridges, gentle hills, and flat valley floors. To our knowledge, no public benchmark exists for DEM-based mountain segmentation, which motivated the construction of a dedicated dataset.
The raw DEM tiles were cropped into 256×256 pixel patches with a 50% overlap stride to increase sample diversity, yielding a total of 3,200 patches. The study area was first divided into geographically disjoint regions for training, validation, and test (7:1:2 ratio); overlapping patches were then extracted only within each region, ensuring that no tile appears in more than one split.
To generate ground-truth labels, we used a semi-automatic pipeline based on multi-scale terrain features. First, key topographic metrics including slope, ruggedness, and

the Multi-scale Topographic Position Index (TPI) were extracted. Using the scale dependence of landforms, these metrics provided a preliminary identification of ridge cores [19]. Multi-directional morphological operations and elevation-constrained dilation were then applied to reconstruct discontinuous ridges, while gradient magnitude constraints preserved natural serrated boundaries [15]. Following the Minimum Mapping Unit principle [1], fragmented patches smaller than 1,000 pixels were merged based on elevation similarity. This morphological refinement was applied only to the ground-truth labels and is *not* part of the model's inference pipeline; the network outputs raw pixel-level predictions without any post-processing. The resulting algorithmic labels were then reviewed and corrected by two annotators with expertise in geomorphology; disagreements were resolved through discussion to produce the final binary segmentation masks (mountain vs. non-mountain). The foreground-to-background ratio across the dataset is approximately 3:7, confirming class imbalance that motivates our use of Dice loss.

4.2 Implementation Details

Experiments were conducted on an NVIDIA RTX 3090 GPU. We used the Adam optimizer with an initial learning rate of 1×10^{-4} and a decay scheduler. Training ran for 100 epochs with dropout of 0.3 and early stopping (patience 10). We evaluated segmentation using the Dice coefficient, Intersection over Union (IoU), Precision, Recall, and F1-Score.

Computational overhead. As shown in Table 1, TAA adds few parameters: the Terrain branch has two lightweight convolution layers and a fixed Sobel kernel, the Elevation branch reuses the standard SE-style MLP bottleneck, and the Spatial branch applies a single 7×7 convolution. The two learnable scalars (λ, γ) are negligible. In total, the TAA overhead is comparable to a single SE block. For HAAM, since SimAM is parameter-free and CA/SE/CBAM are already lightweight, the total parameter increase is small relative to the U-Net backbone.

4.3 Experimental Results with a Single Attention Mechanism

Fig. 3 and Table 2 show that adding a single attention mechanism generally improves performance on terrain segmentation.

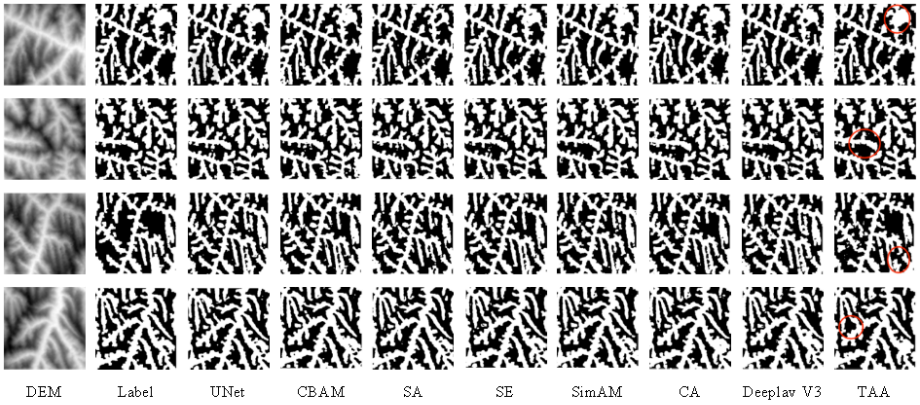


Fig. 3. Segmentation results with attention mechanisms. Red circles highlight the superior details captured by our method.

Performance of the proposed TAA module. TAA achieved the highest performance across all metrics. Compared to the U-Net baseline, TAA improved IoU by 3.3% and Dice by 2.0%. We attribute this gain to the parallel computation and fusion of three terrain-aware branches: elevation, slope, and spatial.

Varying impact of general-purpose mechanisms. Among the general-purpose mechanisms, SE, CA, and SimAM all reached an IoU of 84.5, a modest 0.7-point gain over the baseline. These three modules perform almost identically despite different architectures, suggesting that generic attention has a limited effect on single-channel DEM data. In contrast, the domain-aware TAA reaches 87.1 (+3.3), a larger gain that highlights the value of terrain-specific descriptors. CBAM and SA performed below the baseline, likely because CBAM's serial channel-spatial pipeline introduces redundancy on low-dimensional DEM features, while SA's quadratic cost over-emphasizes global modeling at the expense of local terrain details.

Limitations of generic segmentation models. DeepLabV3+ [2] did not outperform the baseline, suggesting that generic segmentation architectures transfer poorly to DEM data. Its decoder appears inadequate for recovering the weak textures and indistinct boundaries characteristic of DEM data, highlighting the challenges of DEM segmentation.

Table 2. Quantitative comparison of individual attention mechanisms on terrain segmentation tasks.

Model	Dice	IoU	Prec.	Recall	F1
U-Net	91.1	83.8	90.7	91.6	91.1
(+)CBAM [20]	90.9	83.4↓0.4	90.9	91.0	90.9
(+)SA [17]	90.0	82.0↓1.8	89.4	90.8	90.0
(+)SE [7]	91.5	84.5↑0.7	91.4	91.8	91.5
(+)SimAM [24]	91.5	84.5↑0.7	91.5	91.7	91.5
(+)CA [6]	91.5	84.5↑0.7	90.8	92.4	91.5
DeepLabV3+ [2]	90.9	83.0↓0.8	90.0	91.9	90.9
TAA	**93.1**	**87.1↑3.3**	**92.6**	**93.7**	**93.1**

4.4 Results of the Combined Attention Mechanism

Table 3 shows that combination strategy matters and that the proposed HAAM performs best.

Necessity of hierarchical design. As shown in Table 3, effective combinations (e.g., CBAM+SimAM) work because they use the strengths of each module. Simply piling up modules (e.g., Parallel CBAM+SimAM) leads to conflicts and degrades performance, sometimes dropping below the baseline. These results support the HAAM design: instead of naive stacking, it assigns different attention modules to the layers where they work best.

Performance of the TAA-HAAM combination. TAA-HAAM achieved the best performance with an IoU of 88.2, a 4.4% improvement over the baseline. We attribute this to two factors. First, HAAM places modules by depth: SimAM in shallow layers to keep texture details, CA in deep layers to extract semantics. Second, TAA incorporates geographic information (e.g., elevation) into the features. Together, the model captures both local details and global shapes efficiently.

Table 3. Quantitative evaluation of different attention combination strategies and the proposed HAAM architecture.

Model	Dice	IoU	Prec.	Recall	F1
Baseline	91.1	83.8	90.7	91.6	91.1
(+)CBAM+SimAM	91.9	85.1↑1.3	91.1	92.8	91.9
(+)P_CBAM_SimAM	90.4	82.7↓1.1	90.3	90.7	90.4
(+)P_SE_CA	91.6	84.7↑0.9	90.7	92.8	91.6
(+)SE+CBAM	91.2	83.9↑0.1	91.2	91.3	91.2
(+)SimAM+SE	91.3	84.2↑0.4	91.4	91.4	91.3
(+)Triplet_seq	89.3	80.8↓3.0	90.6	88.2	89.3
HAAM	**93.5**	**88.0↑4.2**	**94.0**	**93.2**	**93.5**
HAAM+TAA	**93.8**	**88.2↑4.4**	**94.0**	**93.7**	**93.8**

4.5 Ablation Study

To evaluate each branch and the fusion design, we conducted ablation experiments (Table 4).

Dominance of core features. Single-branch results show that Elevation (85.9) and Terrain Slope (85.4) are the main contributors, outperforming the Spatial branch (75.8). This matches the design: the Elevation branch captures global altitude context, while the Terrain branch extracts geometric boundaries. These two provide the core semantic and structural information for the task.

Effect of the weighting mechanism. The dual-branch combination of "slope + elevation" increased IoU to 86.7, surpassing either branch alone. This supports the use of the balancing coefficient λ in the TAA design. By weighting local gradients (slope) and global context (elevation), the model combines edge detection with semantic localization.

Collaborative gain of the full module. The fully integrated TAA module achieves the highest performance. The Spatial branch alone performs poorly (75.8) because texture statistics carry limited semantic meaning, but it is still important in the full model. As shown in (6), the Spatial map multiplies with the other features rather than adding to them. This allows it to act as a *highlighter*, sharpening details in complex areas for the Elevation and Slope branches and improving the final result.

Table 4. Ablation studies of proposed components.

Terrain	Elev.	Spatial	Dice	IoU	Prec.	Recall	F1
√			92.1	85.4	90.2	94.2	92.1
	√		92.4	85.9	91.0	93.8	92.4
		√	86.2	75.8	82.0	90.9	86.2
√	√		92.9	86.7	92.2	93.6	92.9
	√	√	92.4	85.8	91.4	93.4	92.4
√		√	92.3	85.6	91.3	93.2	92.3
√	√	√	93.1	87.1	92.6	93.7	93.1

5 Conclusions

We presented the TAA module to improve terrain segmentation accuracy. By embedding geographical priors into the network, this module improves feature extraction, as confirmed by ablation experiments. We also introduced HAAM, which assigns different attention mechanisms by depth for multi-scale feature extraction. On our ALOS PALSAR-derived dataset, the combined TAA+HAAM model achieved the highest IoU among the evaluated configurations.

Limitations. Our evaluation uses a single DEM source (ALOS PALSAR, 12.5 m); since no public benchmark exists for DEM mountain segmentation, cross-sensor validation (e.g., SRTM or LiDAR-derived DEMs) is left for future work. The HAAM

placement is guided by complexity analysis and empirical comparison rather than exhaustive search; systematic exploration of placement strategies may yield further gains. Incorporating higher-order terrain features such as curvature and aspect into TAA is also left to future work.

Disclosure of Interests.
The authors have no competing interests to declare that are relevant to the content of this article.

6 References

1. **[1]** Blaschke, T.: Object based image analysis for remote sensing. ISPRS journal of photogrammetry and remote sensing **65**(1), 2–16 (2010)
2. **[2]** Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 801–818 (2018)
3. **[3]** Dai, W., Yang, X., Na, J., Li, J., Brus, D., Xiong, L., Tang, G., Huang, X.: Effects of DEM resolution on the accuracy of gully maps in loess hilly areas. Catena **177**, 114–125 (2019)
4. **[4]** Dai, Z., Liu, H., Le, Q.V., Tan, M.: CoAtNet: Marrying convolution and attention for all data sizes. Advances in neural information processing systems **34**, 3965–3977 (2021)
5. **[5]** Ghaffarian, S., Valente, J., Van Der Voort, M., Tekinerdogan, B.: Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review. Remote Sensing **13**(15), 2965 (2021)
6. **[6]** Hou, Q., Zhou, D., Feng, J.: Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13713–13722 (2021)
7. **[7]** Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7132–7141 (2018)
8. **[8]** Jiao, H., Li, F., Wei, H., Liu, W.: An improved shoulder line extraction method fusing edge detection and regional growing algorithm. Applied Sciences **12**(24), 12662 (2022)
9. **[9]** Jonnala, N.S., Siraaj, S., Prastuti, Y., Chinnababu, P., Praveen babu, B., Bansal, S., Upadhyaya, P., Prakash, K., Faruque, M.R.I., Al-Mugren, K.: AER U-Net: attention-enhanced multi-scale residual U-Net structure for water body segmentation using Sentinel-2 satellite images. Scientific Reports **15**(1), 16099 (2025)
10. **[10]** Li, K., Wang, Y., Gao, P., Song, G., Liu, Y., Li, H., Qiao, Y.: UniFormer: Unified transformer for efficient spatiotemporal representation learning. arXiv preprint arXiv:2201.04676 (2022)
11. **[11]** Liu, Z., Zhang, H., Dong, L., Sun, Z., Wu, S., Zhang, B., Yuan, L., Wang, Z., Jia, Q.: An optimised region-growing algorithm for extraction of the loess shoulder-line from DEMs. ISPRS International Journal of Geo-Information **12**(4), 140 (2023)
12. **[12]** Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. IEEE (2016)
13. **[13]** Misra, D., Nalamada, T., Arasanipalai, A.U., Hou, Q.: Rotate to attend: Convolutional triplet attention module. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 3139–3148 (2021)

14. **[14]** Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
15. **[15]** Straumann, R.K., Purves, R.S.: Delineation of valleys and valley floors. In: International Conference on Geographic Information Science. pp. 320–336. Springer (2008)
16. **[16]** Tang, G.: Progress of digital elevation models and digital terrain analysis in china. Acta Geographica Sinica **69**(9), 1305–1325 (2014), (in Chinese)
17. **[17]** Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
18. **[18]** Wang, H., Li, X.: Expanding horizons: U-Net enhancements for semantic segmentation, forecasting, and super-resolution in ocean remote sensing. Journal of Remote Sensing **4**, 0196 (2024)
19. **[19]** Weiss, A.: Topographic position and landforms analysis. In: Poster presentation, ESRI user conference, San Diego, CA. vol. 200 (2001)
20. **[20]** Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
21. **[21]** Wu, Y., Wang, F., Zhao, P., Zhou, M., Geng, S., Zhang, D.: UNet with multibranch prior information encoding for building segmentation in remote sensing images. Advances in Space Research (2025)
22. **[22]** Yan, G., Tang, G., Chen, J., Li, F., Yang, X., Xiong, L., Lu, D.: Modeling computer sight based on DEM data to detect terrain breaks caused by gully erosion on the loess plateau. Catena **237**, 107837 (2024)
23. **[23]** Yang, J., Xu, J., Lv, Y., Zhou, C., Zhu, Y., Cheng, W.: Deep learning-based automated terrain classification using high-resolution DEM data. International Journal of Applied Earth Observation and Geoinformation **118**, 103249 (2023)
24. **[24]** Yang, L., Zhang, R.Y., Li, L., Xie, X.: SimAM: A simple, parameter-free attention module for convolutional neural networks. In: International conference on machine learning. pp. 11863–11874. PMLR (2021)
25. **[25]** Yu, A., Quan, Y., Yu, R., Guo, W., Wang, X., Hong, D., Zhang, H., Chen, J., Hu, Q., He, P.: Deep learning methods for semantic segmentation in remote sensing with small data: A survey. Remote sensing **15**(20), 4987 (2023)
26. **[26]** Zhang, S., Rao, Q., Wang, L., Tang, T., Chen, C.: KPV-UNet: KAN PP-VSSA UNet for remote image segmentation. Electronics **14**(13), 2534 (2025)