

Article

TRB-Net: Terrain-Residual and Boundary-Assisted Multimodal Martian Landslide Segmentation on a Local MMLSv2 Split

Yu Li ¹, Jinxin He ^{1,*}, Yongzhi Wang ², Ye Zhan ³, Yongbin Yang ¹ and Hanya Zhang ¹¹ College of Earth Sciences, Jilin University, Changchun 130061, China; yuli24@mails.jlu.edu.cn (Y.L.); hanya24@mails.jlu.edu.cn (H.Z.)² College of Geoexploration Science and Technology, Jilin University, Changchun 130012, China; wangyongzhi@jlu.edu.cn³ Aviation Operations Service College, Aviation University Air Force, Changchun 130021, China; zhongjiexu@jlu.edu.cn

* Correspondence: hejx@jlu.edu.cn

Highlights

What are the main findings?

- TRB-Net integrates terrain-residual fusion and boundary-assisted supervision to segment Martian landslides from multimodal orbital remote-sensing data, including image, topographic, and thermophysical information.
- The model achieves competitive segmentation performance while improving the interpretability of landslide recognition through explicit terrain and boundary cues.

What is the implication of the main finding?

- The proposed framework provides an interpretable deep-learning baseline for large-scale Martian landslide mapping and geomorphic feature extraction.
- The results demonstrate the value of combining image texture, terrain information, and boundary supervision for planetary remote-sensing segmentation tasks.

Abstract

Martian landslide segmentation is challenging because annotated samples are limited and landslide deposits can have weak boundaries, heterogeneous textures, and visual similarity to crater rims and canyon walls. This study evaluates a terrain-residual and boundary-assisted network (TRB-Net) on the locally available MMLSv2 train/validation/test split. TRB-Net combines RGB texture with digital elevation model (DEM), slope, thermal inertia, and grayscale information through terrain-residual fusion, an atrous spatial pyramid pooling decoder, and auxiliary boundary supervision. The compact evaluation checkpoint, using a validation-selected threshold of 0.55, achieves an mIoU of 0.8060, foreground IoU of 0.7502, F1-score of 0.8573, precision of 0.8473, and recall of 0.8676 with 5.255 million parameters. In same-split comparisons, DeepLabV3+ obtains the highest overlap scores, while TRB-Net provides competitive segmentation and an explicit architecture for tracing how terrain and boundary cues enter the prediction. These results apply only to the local MMLSv2 split; geographically isolated and large-area Martian mapping performance were not evaluated.

Keywords: Martian landslide; semantic segmentation; multimodal remote sensing; terrain-guided fusion; boundary supervision; MMLSv2



Academic Editor: Giancarlo Bellucci

Received: 29 May 2026

Revised: 15 July 2026

Accepted: 16 July 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)[Attribution \(CC BY\)](#) license.

1. Introduction

Martian landslides are among the most important mass-wasting landforms on Mars and are widely distributed in Valles Marineris, crater walls, and canyon slopes [1–4]. Early geomorphological studies investigated the morphology, structural setting, and possible emplacement mechanisms of large landslides in Valles Marineris [1,2]. Later landslide inventories and high-resolution morphological analyses further provided evidence for their spatial distribution, dynamic implications, and geological significance [3,4]. These studies indicate that Martian landslides are not only products of surface evolution but also key indicators for understanding slope stability, mass transport, and possible volatile-related processes. Figure 1 provides the regional topographic context for the Valles Marineris canyon system.

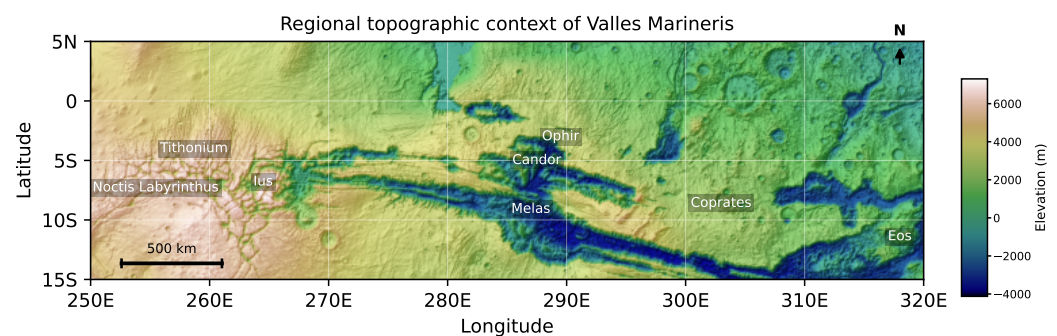


Figure 1. Regional topographic context of Valles Marineris used to introduce the geomorphic setting of Martian landslide segmentation. The map shows MGS MOLA shaded relief and elevation for the study region, with major canyon-system place names, a scale bar, a north arrow, and an elevation color bar. No landslide labels, candidate points, or dataset display panels are shown in this figure; it is used only as a regional geomorphic context map.

Automatic Martian landslide segmentation remains difficult for several reasons. First, Martian landslides exhibit complex morphologies, ranging from large continuous deposits to small, fragmented, and boundary-ambiguous features. Second, Martian observations are derived from heterogeneous sensors and include RGB imagery, digital elevation models (DEM), slope maps, thermal inertia, and grayscale contextual imagery, which differ substantially in spatial resolution, statistical distribution, and physical meaning [5–7]. Third, the number of publicly available annotated samples remains limited, placing the task in a small-data regime [5,6]. Finally, crater rims, valley walls, shadows, and rocky outcrops can show geomorphic patterns similar to landslides, leading to false positives and false negatives.

To address these challenges, this study proposes TRB-Net, a terrain-residual and boundary-assisted multimodal network. Unlike MarsLS-Net, which provides an end-to-end multimodal segmentation framework [5], and DualSwinFusionSeg, which relies on dual Swin Transformer V2 encoders and a UNet++ decoder [7], TRB-Net emphasizes lightweight dual-branch modeling, terrain-residual guided fusion, and explicit boundary-assisted supervision. The key idea is that the RGB branch extracts optical texture and morphological features, whereas the terrain branch extracts DEM, slope, thermal inertia, and grayscale contextual features. Instead of simply concatenating multimodal features, terrain features are injected into RGB-dominant representations as residual enhancement cues. The reported ablations test how this design is associated with overlap and boundary metrics on the local split.

The novelty of this work is therefore not the use of a generic semantic-segmentation backbone alone. It lies in a terrain-aware multimodal design for the local MMLSv2 segmentation task: RGB and terrain-thermophysical channels are encoded separately, terrain cues

are introduced as residual guidance rather than as undifferentiated early-fusion channels, and boundary supervision is used to regularize annotated landslide margins. Ablation, stability, uncertainty, boundary-sensitive, and per-tile distribution analyses are used to relate network behavior to the available terrain channels and mask geometry. These analyses provide diagnostic traceability on the reported split; they are not evidence of cross-region generalization or universally improved interpretability.

To achieve this objective, the study addresses four technical and analytical questions:

- (1) An RGB-terrain dual-branch multimodal encoding framework is proposed for Martian landslide segmentation, enabling separate modeling of optical texture features and terrain-thermophysical features instead of treating all seven channels as a single undifferentiated input stack.
- (2) A terrain-residual guided fusion module is designed to inject terrain constraints such as DEM, slope, thermal inertia, and grayscale context into RGB-dominant features in a residual enhancement manner.
- (3) Boundary-assisted supervision is introduced by automatically generating boundary labels from ground-truth masks and jointly optimizing an auxiliary boundary loss.
- (4) Controlled experiments on the local MMLSv2 split are used to analyze terrain fusion, boundary supervision, multimodal inputs, model stability, and sample-level uncertainty. The evaluation is framed as a same-split design and diagnostic study rather than a claim of broad Martian generalization.

2. Background and Related Work

2.1. Martian Landslide Geomorphology

Early research on Martian landslides mainly focused on geomorphological interpretation of large landslides in Valles Marineris. Lucchitta [1] systematically described landslide morphology and possible transport mechanisms in this region. Quantin et al. [2] further analyzed the geometry and morphology of Valles Marineris landslides and discussed their dynamic implications. Brunetti et al. [3] developed a geomorphological landslide inventory for Valles Marineris, supporting spatial distribution and classification analyses. Mazzanti et al. [4] used morphological evidence to infer high landslide velocities, highlighting the value of high-resolution geomorphological analysis for understanding Martian mass movements. Global mobility analyses further show that Martian landslide behavior is related to location, typology, runout, and possible ice-related conditions [8]. These studies motivate segmentation models that preserve topographic and boundary information rather than relying only on image texture.

2.2. Martian Landslide Datasets and Benchmarks

MarsLS-Net introduced a multimodal benchmark dataset and an end-to-end model for Martian landslide segmentation [5]. MMLSv2 extends the original MMLS benchmark into a seven-channel multimodal dataset containing RGB, DEM, slope, thermal inertia, and grayscale channels, and provides a geographically isolated test set for spatial-generalization evaluation [6]. The physical relevance of these modalities is consistent with planetary remote-sensing practice: orbital cameras provide regional optical context [9], topographic products support relief and slope analysis [10], and thermal infrared observations provide thermophysical constraints [11]. More recently, DualSwinFusionSeg was proposed for multimodal Martian landslide segmentation using dual Swin Transformer V2 encoders, multi-scale fusion, and a UNet++ decoder [7]. The MMLSv2 reference has been updated to its published CVPR Workshops 2026 version; the DualSwinFusionSeg source used here is the arXiv preprint [7].

2.3. Deep Learning-Based Landslide Segmentation

Deep learning has been widely used for semantic segmentation of natural and remote sensing scenes. U-Net introduced the encoder-decoder architecture with skip connections and became a foundational segmentation model [12]. UNet++ improved skip pathways through dense nested connections [13], while DeepLabV3+ incorporated atrous spatial pyramid pooling (ASPP) and a decoder to improve multi-scale context modeling and boundary recovery [14]. Transformer-based models such as SegFormer, Swin Transformer, and Swin Transformer V2 further improved hierarchical visual representation and long-range contextual modeling [15–17]. For landslide segmentation, Landslide4Sense demonstrated the usefulness of multimodal inputs including optical imagery and terrain features [18], while D2FLS-Net showed the value of DEM-guided fusion for landslide segmentation [19]. Recent landslide studies in 2026 further emphasize multi-source data fusion, dual-branch feature modeling, and boundary-aware segmentation for remote-sensing landslide identification [20–22]. Earlier machine-learning work on extraterrestrial landslides also used DEM, slope, thermal inertia, and geological variables to classify Valles Marineris landslides, supporting the use of physically interpretable terrain descriptors [23].

2.4. Multimodal Fusion and Attention Mechanisms

Multimodal fusion is essential when different sensors provide complementary information. Existing approaches commonly use early fusion, feature concatenation, weighted summation, or attention-based fusion. Attention mechanisms such as SE, CBAM, and ECA-Net improve feature representations through channel recalibration, channel-spatial attention, or efficient channel attention [24–26]. These studies motivate the terrain-guided and boundary-aware design of TRB-Net. Different from simple concatenation or heavy Transformer-only fusion, the proposed model injects terrain information as residual enhancement cues into RGB-dominant representations.

3. Dataset and Preprocessing

3.1. MMLSv2 Dataset

Experiments are conducted on the MMLSv2 dataset, a multimodal remote sensing dataset designed for binary semantic segmentation of Martian landslides [6]. Each sample is a 128×128 multiband GeoTIFF tile containing seven co-registered channels, including RGB, DEM, slope, thermal inertia, and grayscale imagery, with a corresponding binary landslide mask. The MMLSv2 dataset used in this study was obtained from the public repository associated with Paheding et al. [6]. The authors did not redefine the train/validation/test partition. Instead, the locally available folders in the downloaded package were used directly, containing 465 training samples, 66 validation samples, and 133 test samples. The geographically isolated test set described by the public MMLSv2 documentation was not available as a separate local folder in the package used for these experiments. Therefore, the reported results correspond to the local MMLSv2 test split rather than an official hidden challenge test set or the geographically isolated test set. Table 1 summarizes the dataset composition used in this study.

This split definition is important for interpreting the results. Because the isolated test subset was not available locally, this study does not claim that the training and test tiles are spatially independent at a regional scale, nor does it claim that spatial leakage has been fully excluded in the sense of a geographically isolated benchmark. To avoid overstating generalization, all comparisons are described as same-split local MMLSv2 experiments. Evaluation on the official isolated test set or on independently prepared Martian regions remains necessary for assessing cross-region transfer.

Table 1. Dataset used in this study [6].

Dataset	Task	Modalities/Channels	Patch Size	Split Used in This Study
MMLSv2	Martian landslide semantic segmentation	RGB, DEM, slope, thermal inertia, and grayscale; seven channels	128 × 128	Provided local split: train 465/validation 66/test 133; isolated-test folder unavailable locally

3.2. Seven-Channel Multimodal Input

The seven-channel input is divided into two modality groups. The RGB branch takes the R, G, and B channels as input to learn optical texture, brightness, shadows, and landslide morphology. The terrain branch takes DEM, thermal inertia, slope, and grayscale channels as input to learn elevation relief, slope breaks, thermophysical differences, and terrain context. This separation is motivated by the different physical meanings and statistical distributions of RGB and terrain-thermophysical data.

Each input source represents different potential evidence for landslide identification. RGB imagery captures surface texture, albedo contrast, shadow patterns, and morphological continuity. DEM represents elevation relief, and slope represents steep scarps, slope breaks, and margin transitions. Thermal inertia describes thermophysical variation, although its effective contribution can be limited by spatial resolution and co-registration uncertainty. The grayscale contextual channel supplies an additional image representation. In the modality-level ablation in Section 7, the multiscale reference scores higher than the RGB-only and terrain-only variants, while the individual channel-removal experiment contains a stated no-DEM exception. These are split-specific experimental observations rather than evidence that any channel has a universally fixed geological contribution.

3.3. Data Normalization and Augmentation

The preprocessing pipeline includes GeoTIFF reading, channel rearrangement, normalization, binary mask conversion, and spatial resizing. During training, geometric augmentations such as horizontal flipping, vertical flipping, and rotation are applied. For physically meaningful channels such as DEM, slope, and thermal inertia, strong photometric perturbations are avoided to prevent distortion of terrain and thermophysical information. Photometric enhancement is limited to RGB and grayscale channels when used.

4. Methodology

4.1. Overall Architecture of TRB-Net

TRB-Net consists of an RGB branch, a terrain branch, terrain-residual guided fusion modules, an ASPP multi-scale context module, a U-shaped decoder, a segmentation head, and a boundary head. The seven-channel input is first divided into RGB and terrain inputs. The two branches extract multi-scale features separately. Corresponding features are fused by terrain-residual guided fusion modules, and the fused high-level features are enhanced by the ASPP module for multi-scale contextual representation. The decoder then outputs both a landslide segmentation mask and an auxiliary boundary map.

Figure 2 is organized into four functional parts. The input part separates optical RGB channels from terrain-thermophysical channels. The encoder part extracts modality-specific features at multiple scales. The terrain-residual fusion part injects terrain guidance into RGB-dominant features before ASPP-based multi-scale context modeling. The output part contains the main segmentation head and the auxiliary boundary head, which are optimized jointly during training.

Two implementation configurations are reported and are kept separate throughout the manuscript. TRB-Net-Multiscale follows the conceptual diagram directly by maintaining

modality-specific features and applying terrain-residual fusion at four encoder scales; it is used only for the component ablations. TRB-Net-Compact, used for the principal same-split comparison, retains separate RGB and four-channel terrain stems and applies terrain-residual fusion once at the low-level stem, followed by a shared convolutional encoder and a lightweight ASPP decoder. The configurations also use different training objectives and validation-selected thresholds, as specified in Table 2.

TRB-Net architecture

Explicit terrain-residual fusion and boundary-assisted supervision for interpretable Martian landslide segmentation

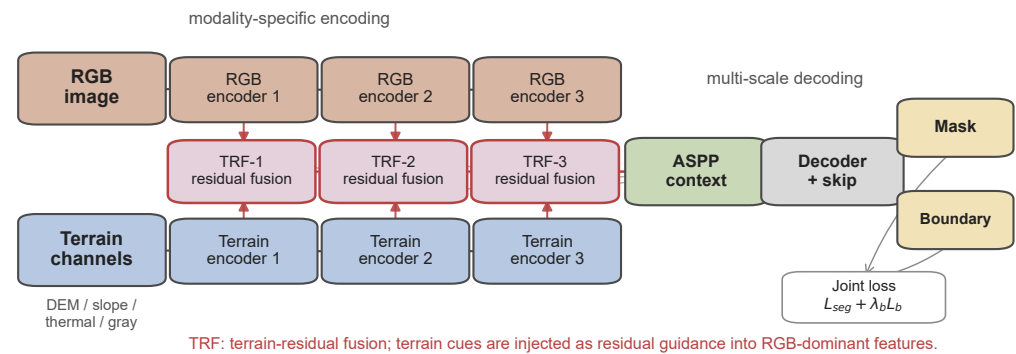


Figure 2. Conceptual overview of the TRB-Net design. Brown blocks denote the RGB pathway, blue blocks denote the terrain-thermophysical pathway, red blocks denote terrain-residual fusion (TRF), the green block denotes ASPP context aggregation, the gray block denotes the decoder and skip connections, and yellow blocks denote the mask and boundary outputs. Horizontal lines indicate forward feature propagation, vertical red arrows indicate injection of terrain guidance into RGB-dominant features, and the two output branches contribute to the joint loss. The multiscale implementation used for component ablations applies TRF at four encoder scales; the compact evaluation implementation applies the same residual-fusion principle at the low-level stem before an ASPP decoder.

Table 2. Experimental settings for the two explicitly distinguished TRB-Net implementations.

Parameter	TRB-Net-Compact	TRB-Net-Multiscale
Role in manuscript	Main comparison	Component ablations
Input size	128×128	128×128
Base channels	48	32
Parameters	5.255 M	6.952 M
Epochs/batch size	220/4	220/4
Optimizer	AdamW	AdamW
Learning rate/weight decay	$1 \times 10^{-4}/5 \times 10^{-4}$	$1 \times 10^{-4}/5 \times 10^{-4}$
Scheduler	Cosine annealing	Cosine annealing
Segmentation loss	weighted BCE + Dice loss	CE + Dice loss + focal loss
Boundary-loss weight	0.10	0.50
Validation-selected threshold	0.55	0.65
Device/mixed precision	RTX 5060 Laptop GPU/enabled	RTX 5060 Laptop GPU/enabled

4.2. Dual-Branch Multimodal Encoder

The RGB branch extracts Martian surface texture, brightness, and morphological features, while the terrain branch extracts elevation, slope, thermal inertia, and grayscale contextual features. A dual-branch encoder is adopted because RGB and terrain-thermophysical data differ substantially in dynamic range, spatial frequency, and pixel semantics. Direct concatenation at shallow layers is used as an alternative in the fusion ablation. DualSwinFusionSeg similarly adopts modality-specific feature extraction for Martian landslide segmentation [7].

4.3. Terrain-Residual Guided Fusion

Terrain-residual guided fusion is the core operation of TRB-Net. At scale i , the RGB feature and terrain feature are denoted by $F_{\text{RGB}}^{(i)}$ and $F_{\text{ter}}^{(i)}$, respectively. The transform $G_i(\cdot)$ projects and gates the terrain feature to the RGB feature dimension. The fused feature is defined as

$$F_{\text{fused}}^{(i)} = F_{\text{RGB}}^{(i)} + \alpha_i G_i(F_{\text{ter}}^{(i)}). \quad (1)$$

Here, $F_{\text{fused}}^{(i)}$ is the fused output and α_i is a learnable scalar controlling the terrain-residual contribution. Unlike the concatenation plus 1×1 convolution strategy used in DualSwinFusionSeg [7], TRB-Net injects terrain information into RGB-dominant features as residual enhancement cues. This mechanism is inspired by DEM-guided fusion [19], but it emphasizes residual terrain constraints on RGB texture features rather than relying entirely on cross-attention modules.

4.4. Multi-Scale Context Enhancement

Martian landslides show strong scale variation, including large continuous deposits and small fragmented patches. TRB-Net introduces an ASPP/multi-scale context module between the encoder and decoder to enlarge the effective receptive field and capture multi-scale contextual information. This design follows the idea of multi-receptive-field context encoding in DeepLabV3+ [14].

4.5. Boundary-Assisted Supervision

Landslide boundaries often correspond to key geomorphic units such as head scarps, lateral shear margins, and depositional fronts. Models such as U-Net and UNet++ mainly recover spatial details through skip connections [12,13], whereas TRB-Net further introduces an explicit boundary-assisted branch. Boundary labels are generated from the ground-truth mask using morphological dilation and erosion:

$$B = \text{Dilate}_{3 \times 3}(M) - \text{Erode}_{3 \times 3}(M). \quad (2)$$

Here, $M \in \{0,1\}^{H \times W}$ is the binary ground-truth mask and B is its one-pixel morphological-gradient boundary target. Two implementations were retained in the reported experiments and are distinguished explicitly. For the compact evaluation checkpoint, the objective is

$$\mathcal{L}_{\text{compact}} = 0.5\mathcal{L}_{\text{WBCE}} + 0.5\mathcal{L}_{\text{Dice}} + 0.1(\mathcal{L}_{\text{BCE}}^b + \mathcal{L}_{\text{Dice}}^b), \quad (3)$$

where $\mathcal{L}_{\text{WBCE}}$ is weighted binary cross-entropy with foreground weight 2.5, $\mathcal{L}_{\text{Dice}}$ is Dice loss, and the superscript b denotes the auxiliary boundary terms. The multiscale configuration used for the component ablations is optimized with

$$\mathcal{L}_{\text{multi}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{Dice}} + 0.4\mathcal{L}_{\text{Focal}} + \lambda_b \mathcal{L}_{\text{BCE}}^b, \quad (4)$$

where $\lambda_b = 0.5$ for the multiscale reference configuration and $\lambda_b = 0$ only in the no-boundary ablation. Dice loss addresses foreground–background imbalance [27], while focal loss emphasizes hard examples [28]. Because the two objectives differ, their reported loss values are configuration-specific and are not compared directly.

4.6. Loss Function and Inference

The compact implementation produces one foreground logit and applies a sigmoid function, whereas the multiscale implementation produces two class logits and applies a

softmax function. In both cases, the foreground probability is converted into a binary mask using the threshold that maximizes foreground IoU on the corresponding validation split. The selected threshold is 0.55 for the compact checkpoint used in the main comparison and 0.65 for the multiscale reference used in the subsequent component-ablation tables. Each ablation variant uses its own validation-selected threshold, as reported with the corresponding results.

5. Experiments

5.1. Experimental Settings

The models were implemented using the PyTorch framework (Meta Platforms, Inc., Menlo Park, CA, USA) and trained on an NVIDIA GeForce RTX 5060 Laptop GPU (NVIDIA Corporation, Santa Clara, CA, USA). Training used the AdamW optimizer, cosine annealing learning-rate scheduling, and mixed-precision computation. AdamW decouples weight decay from gradient-based parameter updates [29]. Table 2 distinguishes the compact checkpoint used in the main comparison from the multiscale configuration used for component ablations.

5.2. Evaluation Metrics

The primary metrics are computed globally: predictions and ground-truth labels from all 133 test tiles are accumulated into one pixel-level confusion matrix. Let TP , TN , FP , and FN denote foreground true positives, background true negatives, false positives, and false negatives, respectively. The metrics are

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad \text{IoU}_{\text{fg}} = \frac{TP}{TP + FP + FN}, \quad (5)$$

$$\text{IoU}_{\text{bg}} = \frac{TN}{TN + FP + FN}, \quad \text{mIoU} = \frac{\text{IoU}_{\text{fg}} + \text{IoU}_{\text{bg}}}{2}, \quad (6)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad \text{F1} = \frac{2 \text{Precision Recall}}{\text{Precision} + \text{Recall}}. \quad (7)$$

A numerical constant of 10^{-7} is added to denominators in the implementation. None of the 133 local test masks is empty (the minimum ground-truth foreground ratio is 0.00098), so no empty-mask special case affects the reported global metrics. The later figures reporting per-tile distributions are explicitly labeled as per-tile analyses and use the same formulas independently for each tile. Foreground IoU directly measures overlap of the landslide class; mIoU averages foreground and background IoU [30].

For the boundary diagnostics, a 3×3 morphological gradient extracts predicted and ground-truth contours. Boundary precision and boundary recall count contour pixels matched within a two-pixel dilation tolerance; boundary F1 is their harmonic mean. Boundary IoU is the intersection divided by the union of the two tolerance-dilated contour bands. Counts are pooled within each evaluation batch (batch size 8), and batch scores are averaged with weights equal to the number of tiles in the batch. These boundary metrics are reported only as diagnostic complements to the globally pooled overlap metrics.

6. Results

6.1. Quantitative Results

TRB-Net-Compact achieves an mIoU of 0.8060, foreground IoU of 0.7502, and F1-score of 0.8573 on the local MMLSv2 test split. Precision and recall are 0.8473 and 0.8676, respectively. The test loss of 0.4883 belongs to the compact BCE–Dice objective in Equation (3); it is not compared with loss values from the multiscale ablation objective. The selected

threshold is 0.55, and the model contains 5.255 million parameters. Table 3 reports the complete compact-checkpoint metric set.

Table 3. Quantitative results of TRB-Net-Compact on the local MMLSv2 test split.

Metric	Value
Loss	0.4883
Accuracy	0.9023
mIoU	0.8060
Foreground IoU	0.7502
F1-score	0.8573
Precision	0.8473
Recall	0.8676
Threshold	0.55
Parameters	5.255 M

6.2. Comparison with Existing Methods

The same-split comparison shows that DeepLabV3+ obtains the highest numerical overlap scores. One possible interpretation is that its convolutional encoder, ASPP context aggregation, and decoder form a strong inductive bias for the 128×128 tiles; the present experiments do not isolate regularization as a causal explanation. TRB-Net-Compact outperforms the local SegFormer-lite and DualSwin-style lite implementations and remains close to the MarsLS-Net-style approximation. Its contribution is therefore restricted to the explicit terrain-residual and boundary-assisted design and the associated diagnostic analyses, not numerical superiority over DeepLabV3+ or all tested baselines.

This comparison clarifies architecture attributes and performance–complexity trade-offs rather than assigning a subjective interpretability score. DeepLabV3+ provides the highest overlap scores, whereas TRB-Net-Compact exposes a separate terrain stem, residual terrain fusion, and an auxiliary boundary output. These attributes make the input pathway traceable but do not, by themselves, prove improved interpretability. The MarsLS-Net-style, DualSwin-style lite, and SegFormer-lite rows are local same-split approximations or lightweight implementations, not official reproduction results from the original authors. Figure 3 summarizes the same-split performance and model-size comparison.

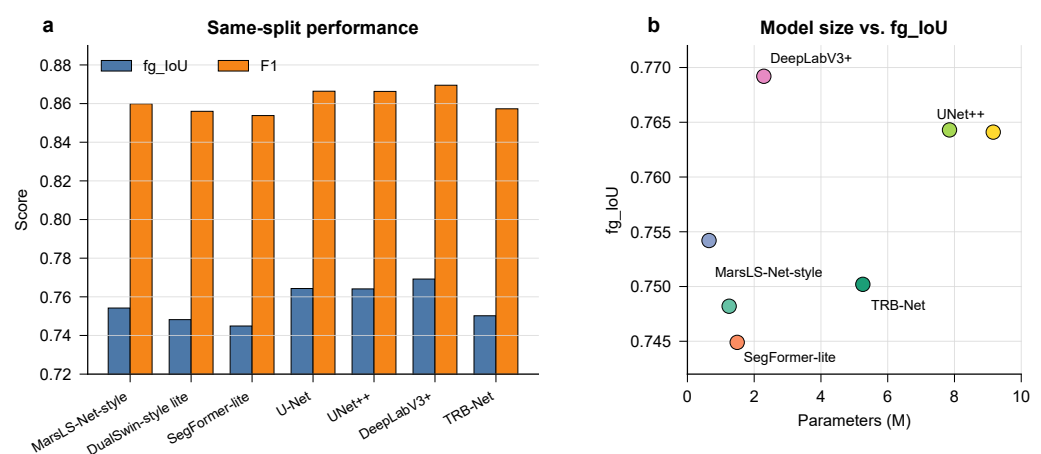


Figure 3. Same-split comparison on the local MMLSv2 test split. (a) Test foreground IoU and F1-score of TRB-Net-Compact and baseline models. (b) Model size versus test foreground IoU, where parameter count is used only as a complexity indicator; the purple point denotes DeepLabV3+.

The subsequent three-seed comparison limits overinterpretation of a single-seed difference on the small local split.

6.3. Qualitative Visualization

Figures 4 and 5 show manually selected illustrative test cases spanning larger continuous masks, smaller or fragmented masks, and visible boundary errors. They were not selected randomly and are not used to estimate model performance. Figure 6 is intentionally a favorable-case diagnostic: it shows the three test tiles with the largest positive per-tile foreground-IoU difference for TRB-Net-Compact relative to DeepLabV3+ (tiles 36, 73, and 0). It therefore demonstrates only that the two models make different local errors. The aggregate results in Table 4, where DeepLabV3+ scores higher, remain the basis for model comparison.

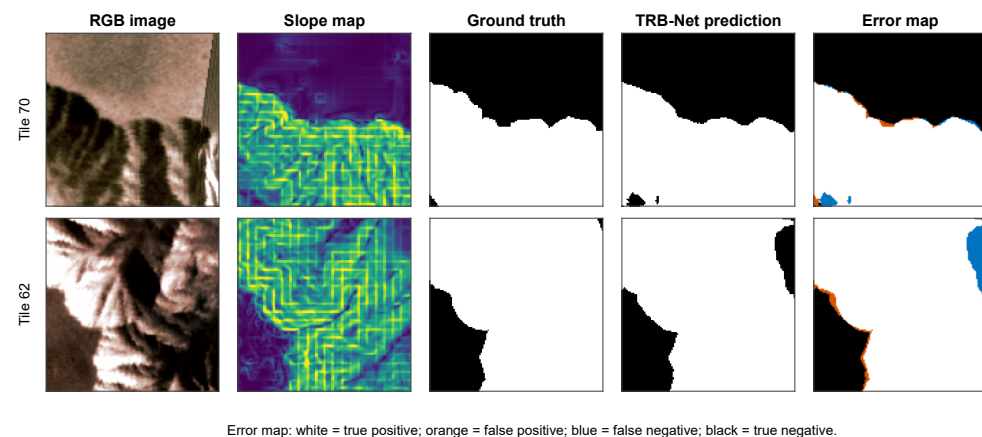


Figure 4. Manually selected illustrative local MMLSv2 test cases, part 1: tile 70 (sample co100039_row00001) and tile 62 (sample co100035_row00018). From left to right, each row shows the RGB image, slope map, ground-truth mask, TRB-Net-Compact prediction, and pixel-level error map. Slope maps use a purple-to-yellow scale from lower to higher slope values. White, orange, blue, and black denote true positives, false positives, false negatives, and true negatives, respectively. These examples were selected to display mask-extent and boundary-error variation rather than by a random or best-score rule.

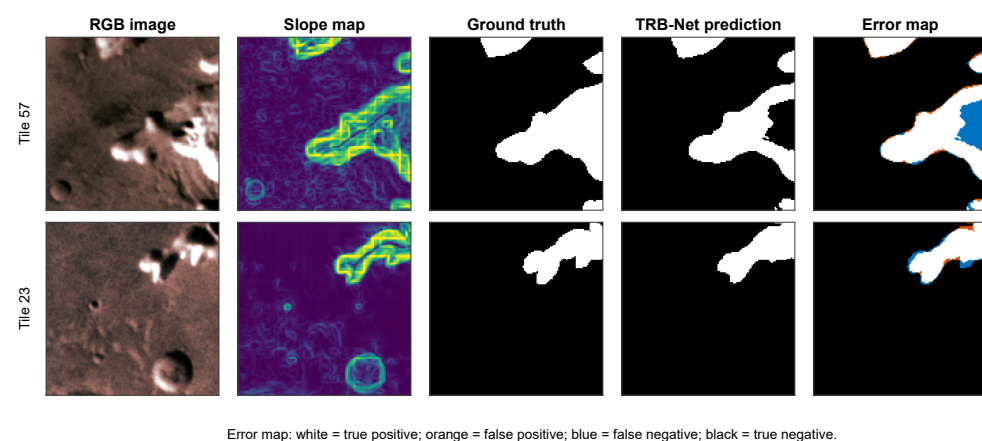


Figure 5. Additional manually selected illustrative local MMLSv2 test cases, part 2: tile 57 (sample co100027_row00014) and tile 23 (sample co100013_row00012). The panel order, color convention, and selection purpose are the same as in Figure 4.

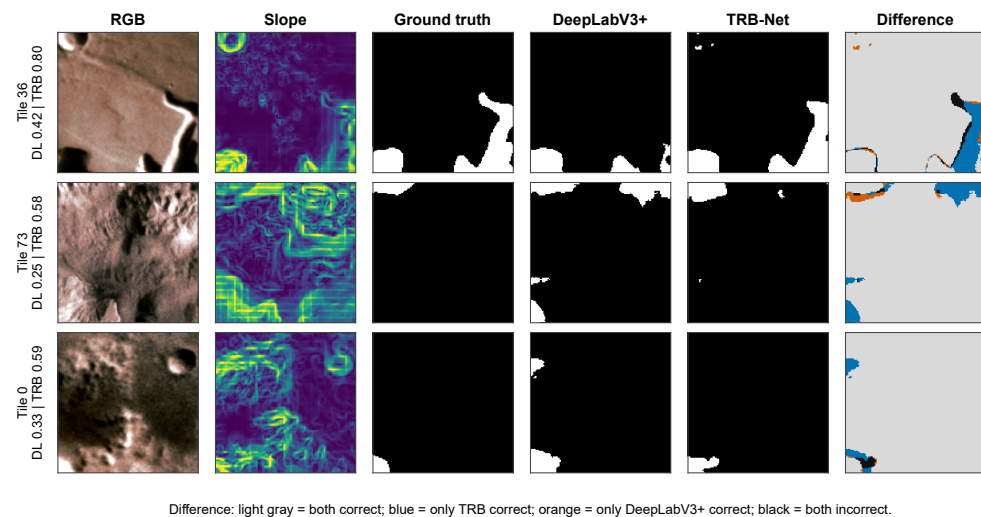


Figure 6. Purposefully selected favorable-case comparison between DeepLabV3+ and TRB-Net-Compact. The three rows are the test tiles with the largest positive TRB-Net-minus-DeepLabV3+ foreground-IoU differences (tiles 36, 73, and 0). Slope maps use purple for lower and yellow for higher values. In the ground-truth and prediction masks, white denotes foreground and black denotes background. In the difference map, blue denotes pixels correctly classified only by TRB-Net-Compact, orange denotes pixels correctly classified only by DeepLabV3+, light gray denotes pixels correctly classified by both, and black denotes pixels misclassified by both. This figure is diagnostic and does not represent aggregate superiority.

Table 4. Same-split comparison on the local MMLSv2 test split. “Style” and “lite” models are local architectural approximations or lightweight implementations and are not official implementations from the cited authors.

Model	Dataset/Split	Parameters (M)	mIoU	Foreground IoU	F1-Score	Precision	Recall
MarsLS-Net-style	MMLSv2 local test	0.647	0.8093	0.7542	0.8599	0.8512	0.8688
DualSwin-style lite	MMLSv2 local test	1.250	0.8047	0.7482	0.8560	0.8484	0.8637
SegFormer-lite	MMLSv2 local test	1.491	0.8002	0.7449	0.8538	0.8311	0.8778
U-Net	MMLSv2 local test	7.851	0.8178	0.7643	0.8664	0.8622	0.8706
UNet++	MMLSv2 local test	9.161	0.8174	0.7641	0.8663	0.8605	0.8721
DeepLabV3+	MMLSv2 local test	2.289	0.8216	0.7692	0.8695	0.8653	0.8738
TRB-Net-Compact	MMLSv2 local test	5.255	0.8060	0.7502	0.8573	0.8473	0.8676

Table 5 summarizes the reported overlap scores and the explicit architecture attributes of DeepLabV3+ and TRB-Net-Compact.

Table 5. Reported overlap scores and explicit architecture attributes; the last three columns are design descriptors, not measured interpretability scores.

Model	mIoU	Foreground IoU	F1-Score	Parameters (M)	Separate Terrain Branch	Residual Terrain Fusion	Boundary Auxiliary Loss
DeepLabV3+	0.8216	0.7692	0.8695	2.289	No	No	No
TRB-Net-Compact	0.8060	0.7502	0.8573	5.255	Yes	Yes	Yes

Table 6 reports the three-seed stability comparison used to limit interpretation of single-seed differences.

Table 6. Three-seed stability comparison on the local MMLSv2 test split. Values are mean $\pm$ standard deviation; each model used the same validation-selected threshold for all three seeds.

Model	Seeds	mIoU	Foreground IoU	F1-Score	Precision	Recall	Threshold
DeepLabV3+	3	0.8173 $\pm$ 0.0045	0.7628 $\pm$ 0.0060	0.8654 $\pm$ 0.0039	0.8688 $\pm$ 0.0090	0.8621 $\pm$ 0.0114	0.60
TRB-Net-Compact	3	0.8116 $\pm$ 0.0056	0.7559 $\pm$ 0.0057	0.8610 $\pm$ 0.0037	0.8607 $\pm$ 0.0136	0.8614 $\pm$ 0.0063	0.55

6.4. Error Analysis

Visual inspection of the displayed test cases suggests that false positives frequently occur near crater rims, valley walls, steep slopes, and rocky outcrops. These landforms can exhibit slope breaks, shadows, and textures that resemble parts of the annotated masks. This description is qualitative because the error categories were not manually annotated across all test tiles. In the single-run terrain-only ablation, precision is 0.3382, which quantitatively shows extensive foreground overprediction when RGB channels are removed.

Visual inspection also suggests that false negatives are concentrated in low-contrast, boundary-ambiguous, small, or fragmented foreground regions. The per-tile and size-stratified analyses quantify the associated performance variation, while the modality ablation shows lower scores for RGB-only input and substantial overprediction for terrain-only input. These observations support complementary use of the two modality groups within the tested configuration without establishing a causal geomorphic mechanism.

Figure 7 further visualizes model uncertainty and pixel-level errors in the three deliberately selected high-error-entropy tiles. For each pixel, uncertainty is computed as binary predictive entropy,

$$U(p) = -p \log(p) - (1 - p) \log(1 - p), \quad (8)$$

where p is the predicted foreground landslide probability. Visual inspection of these selected tiles shows high-entropy responses near predicted margins and weak transitions. Because the tiles were selected by high error-pixel entropy and the error types were not exhaustively annotated, Figure 7 is used only as a diagnostic illustration and does not establish the frequency or geological cause of each error category.

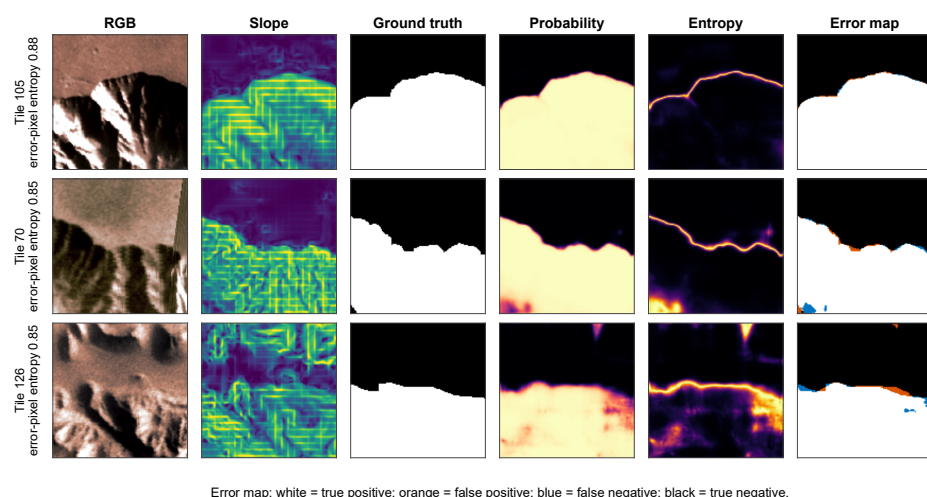


Figure 7. Probability, uncertainty, and error analysis of TRB-Net-Compact on three purposefully selected local MMLSv2 test tiles (tiles 105, 70, and 126). These tiles have the highest mean predictive entropy over their misclassified pixels and therefore illustrate difficult cases rather than a random sample. Slope maps use purple for lower and yellow for higher values; probability and entropy maps use dark colors for lower and bright yellow/white for higher values. The probability map shows foreground probability, and the uncertainty map is binary predictive entropy, $U = -p \log(p) - (1 - p) \log(1 - p)$. Error maps use white for true positives, orange for false positives, blue for false negatives, and black for true negatives.

The distributional analysis in Figure 8 shows that all models exhibit wide per-tile variation despite relatively close aggregate scores. The purpose of this analysis is not to claim per-tile superiority of TRB-Net, but to show that aggregate scores hide substantial sample-level variability. The size-stratified analysis in Figure 9 shows a narrower, higher foreground-IoU distribution in the large-foreground tertile than in the small-foreground tertile. This is an association with ground-truth foreground extent on the local split; boundary-to-area ratio was not independently measured, and no causal attribution to object size or architecture is made.

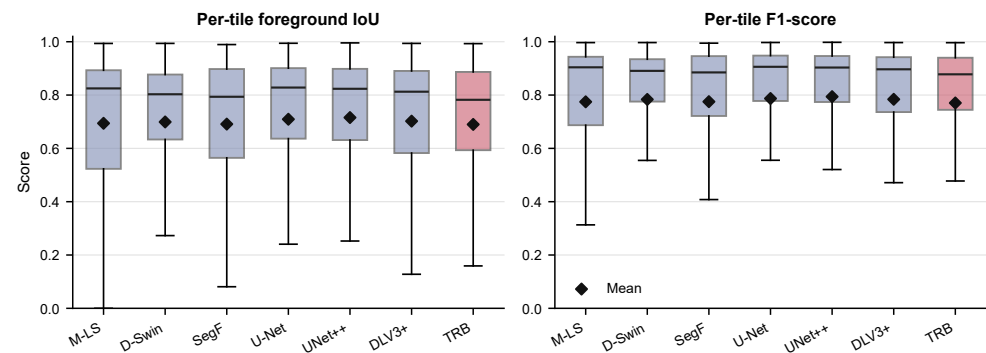


Figure 8. Per-tile performance distribution on the local MMLSv2 test split. Box plots report sample-level foreground IoU and F1-score across the 133 test tiles; boxes indicate the interquartile range, center lines indicate medians, whiskers indicate the non-outlier range, and black diamonds denote mean values. The broad interquartile ranges indicate that local MMLSv2 performance is strongly affected by tile-level heterogeneity rather than by average model score alone. Model names are abbreviated on the horizontal axis for readability.

Finally, Figure 10 quantifies the relationship between uncertainty and error. The Pearson correlation between mean predictive uncertainty and pixel error rate is 0.69 on the local test split. This positive association makes entropy a candidate diagnostic for prioritizing difficult tiles, but it is not a calibration evaluation. In the deliberately selected examples in Figure 7, high-entropy responses are visually concentrated near several mask transitions; the frequency of this pattern was not separately annotated.

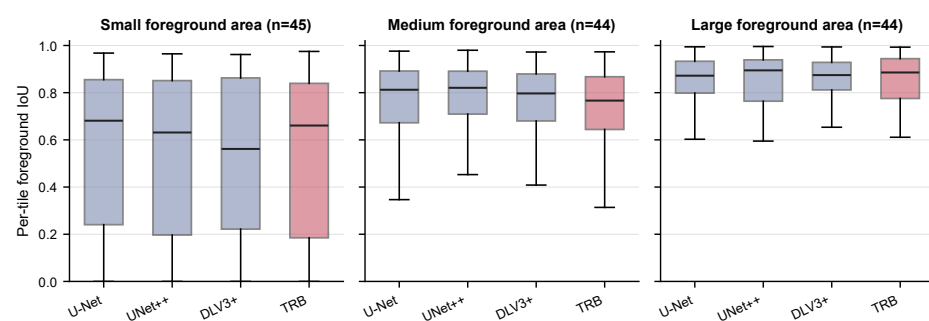


Figure 9. Size-stratified per-tile foreground IoU for representative strong baselines and TRB-Net-Compact on the local MMLSv2 test split. Test tiles are divided by the tertiles of the ground-truth foreground ratio: small, $r \leq 0.183899$ ($n = 45$); medium, $0.183899 < r \leq 0.462769$ ($n = 44$); and large, $r > 0.462769$ ($n = 44$). Model names are abbreviated on the horizontal axis for readability.

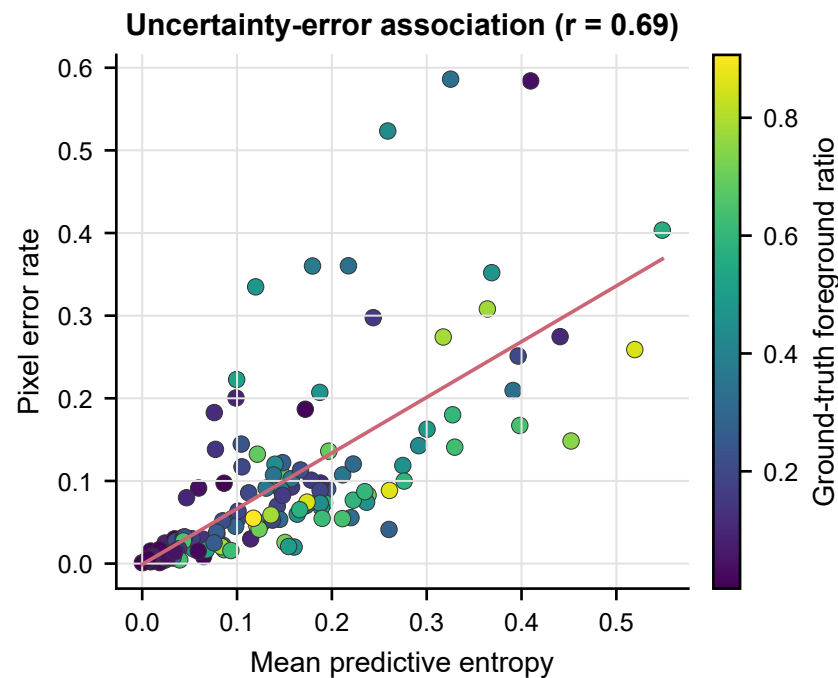


Figure 10. Relationship between mean predictive uncertainty and pixel error rate for TRB-Net-Compact on the local MMLSv2 test split. Each point represents one test tile, mean uncertainty is computed from binary predictive entropy, and point color denotes the ground-truth foreground ratio. The pink line is the ordinary least-squares linear regression trend.

7. Ablation Study

The component ablations use TRB-Net-Multiscale (base channels 32; 6.952 million parameters), not the compact base-channel-48 checkpoint used in the main baseline comparison. This configuration applies terrain-residual fusion at four encoder scales and uses the objective in Equation (4). The experiments isolate input modalities, individual channels, fusion strategy, ASPP, validation-selected threshold, and boundary supervision. All variants were trained for 220 epochs with a batch size of 4. Their results should be compared within this ablation series and not treated as additional runs of TRB-Net-Compact.

7.1. Effect of Terrain Branch

The RGB-only variant achieves an mIoU of 0.6964, foreground IoU of 0.6170, and F1-score of 0.7632. Relative to the multiscale reference in this single-run ablation, removing the four terrain-related channels lowers mIoU by 0.1057, foreground IoU by 0.1303, and F1-score by 0.0922.

The terrain-only variant obtains a recall of 0.9992 but a precision of 0.3382 and mIoU of 0.1694, indicating extensive foreground overprediction. Within the three variants in Table 7, the multiscale reference achieves the highest overlap scores. The result is consistent with complementary information in RGB and terrain channels on this split, but it does not establish the physical cause of the difference.

Table 7. Modality ablation results.

Variant	Input Channels	mIoU	Foreground IoU	F1-Score	Precision	Recall	Threshold
Multiscale reference	RGB + DEM + thermal inertia + slope + grayscale	0.8021	0.7473	0.8554	0.8326	0.8795	0.65
RGB only	RGB	0.6964	0.6170	0.7632	0.7428	0.7847	0.55
Terrain only	DEM + thermal inertia + slope + grayscale	0.1694	0.3381	0.5054	0.3382	0.9992	0.40

7.2. Single-Channel Removal Ablation

Table 8 reports the single-channel removal experiment. Removing grayscale produces the largest score reduction: mIoU decreases from 0.8021 to 0.7687, foreground IoU from 0.7473 to 0.6998, F1-score from 0.8554 to 0.8234, and recall from 0.8795 to 0.8124. Precision changes from 0.8326 to 0.8347. The observed pattern is consistent with the grayscale channel providing additional image contrast, but the experiment does not identify a unique causal mechanism.

Table 8. Single-channel removal ablation results.

Variant	Removed Channel	mIoU	Foreground IoU	F1-Score	Precision	Recall	Threshold
Multiscale reference	None	0.8021	0.7473	0.8554	0.8326	0.8795	0.65
Without DEM	DEM	0.8071	0.7514	0.8580	0.8498	0.8665	0.45
Without slope	slope	0.7964	0.7382	0.8494	0.8396	0.8595	0.70
Without thermal inertia	thermal inertia	0.7915	0.7338	0.8465	0.8262	0.8678	0.55
Without grayscale	grayscale	0.7687	0.6998	0.8234	0.8347	0.8124	0.60

Removing thermal inertia or slope also lowers the overlap scores relative to the multi-scale reference. Removing DEM instead raises mIoU from 0.8021 to 0.8071 and foreground IoU from 0.7473 to 0.7514. Redundancy with slope or split-specific elevation variation are possible interpretations of the DEM result, not established causes. The evidence supports only the reported ranking of these single-run variants on the local split.

7.3. Fusion Strategy Comparison

In the reported single-run fusion ablation in Table 9, terrain-residual fusion achieves the highest mIoU (0.8021), foreground IoU (0.7473), F1-score (0.8554), and recall (0.8795) among the five tested variants. This result is evidence for the selected fusion design within this configuration and local split; it is not a general causal claim that residual fusion is superior to all concatenation or attention mechanisms.

Table 9. Fusion-strategy ablation.

Variant	Fusion Strategy	mIoU	Foreground IoU	F1-Score	Precision	Recall	Threshold
Simple concatenation	$[F_{\text{RGB}}; F_{\text{ter}}]$	0.7845	0.7249	0.8405	0.8212	0.8607	0.50
Sum fusion	$F_{\text{RGB}} + F_{\text{ter}}$	0.7919	0.7334	0.8462	0.8308	0.8622	0.55
Weighted sum	$w_1 F_{\text{RGB}} + w_2 F_{\text{ter}}$	0.7818	0.7204	0.8375	0.8248	0.8505	0.55
Attention fusion	$F_{\text{RGB}} + G(F_{\text{ter}})$	0.7829	0.7182	0.8360	0.8455	0.8266	0.60
Terrain-residual fusion	$F_{\text{RGB}} + \alpha G(F_{\text{ter}})$	0.8021	0.7473	0.8554	0.8326	0.8795	0.65

7.4. ASPP Ablation and Threshold Selection

In the reported single-run ablation in Table 10, removing ASPP lowers test mIoU, foreground IoU, and F1-score to 0.7764, 0.7123, and 0.8320, respectively, compared with 0.8021, 0.7473, and 0.8554 for the multiscale reference. Recall changes from 0.8368 to 0.8795. These values show an association between the tested ASPP configuration and higher overlap scores without isolating scale variation as a causal explanation.

ASPP enlarges the effective receptive field through atrous convolutions with different dilation rates. Its higher score in this ablation is consistent with the need to represent differently sized masks, but no size-specific ASPP effect was separately tested.

Table 10. ASPP ablation.

Variant	ASPP Module	mIoU	Foreground IoU	F1-Score	Precision	Recall	Threshold
Without ASPP	No	0.7764	0.7123	0.8320	0.8272	0.8368	0.55
Multiscale reference	Yes	0.8021	0.7473	0.8554	0.8326	0.8795	0.65

7.5. Validation and Test Performance Under the Optimal Threshold

For TRB-Net-Multiscale, 0.65 maximizes foreground IoU on the validation split and is then applied unchanged to the test split. At this threshold, test mIoU, foreground IoU, and F1-score are 0.8021, 0.7473, and 0.8554. This threshold is specific to the multiscale configuration; the compact checkpoint uses 0.55. No test-set tuning is performed. Table 11 reports the validation and test values at the selected threshold.

Table 11. Performance of TRB-Net-Multiscale at the validation-selected threshold of 0.65.

Dataset	Threshold	Accuracy	mIoU	Foreground IoU	F1-Score	Precision	Recall
Validation	0.65	0.9027	0.8046	0.7453	0.8541	0.8100	0.9032
Test	0.65	0.8994	0.8021	0.7473	0.8554	0.8326	0.8795

7.6. Effect of Boundary Supervision

Without boundary supervision, mIoU, foreground IoU, and F1-score are 0.7955, 0.7352, and 0.8474. With boundary supervision, the corresponding single-run values are 0.8021, 0.7473, and 0.8554. Recall increases from 0.8457 to 0.8795, while precision decreases from 0.8492 to 0.8326. The result shows the observed precision–recall trade-off for this pair of runs; it does not prove that the auxiliary loss will improve every checkpoint. Table 12 reports these values.

Table 12. Boundary supervision ablation results.

Variant	Boundary Weight	mIoU	Foreground IoU	F1-Score	Precision	Recall	Threshold
Multiscale reference	0.50	0.8021	0.7473	0.8554	0.8326	0.8795	0.65
Without boundary supervision	0.00	0.7955	0.7352	0.8474	0.8492	0.8457	0.65

To further examine whether the auxiliary boundary loss changes contour behavior, boundary-sensitive metrics were computed with a two-pixel tolerance around the ground-truth boundary. As shown in Table 13, the run with boundary-assisted supervision has modestly higher boundary F1 and boundary IoU than the paired no-boundary run. This single-run difference is consistent with a contour-regularization effect but does not establish a universal effect across checkpoints or datasets.

Table 13. Boundary-sensitive diagnostic metrics for boundary supervision ablation.

Variant	Boundary F1	Boundary IoU
Multiscale reference	0.5236	0.3234
Without boundary supervision	0.5147	0.3168

7.7. Summary of Ablation Results

Across the reported single-run multiscale ablations, the reference configuration scores higher than RGB-only, terrain-only, alternative fusion, no-ASPP, and no-boundary variants on the principal overlap metrics. The exception is the no-DEM variant, which slightly exceeds the reference. These results identify which tested design choices are associated with the reported scores on the local split; they do not establish universal module effects.

8. Discussion

8.1. Geomorphological Interpretation of Model Inputs

Large Martian landslides can contain head scarps, lateral margins, displaced masses, and depositional fronts [1–4]. RGB channels encode appearance and texture, whereas DEM and slope encode relief and slope transitions; thermal inertia and grayscale provide additional thermophysical and image-context information. TRB-Net makes this assumed division explicit through separate RGB and terrain pathways. The design is therefore traceable in terms of input routing, but the present experiments do not measure human interpretability or identify a new geological process.

The modality ablation provides same-split evidence for this interpretation. Relative to the multiscale reference, RGB-only input lowers mIoU from 0.8021 to 0.6964, while terrain-only input produces recall of 0.9992 and precision of 0.3382. Boundary supervision produces modest changes in both overlap and boundary diagnostics. These observations show how the tested channels and losses affect predictions on the local split; geological statements are limited to consistency with known landform descriptors.

8.2. Comparison with Recent Multimodal Martian Landslide Segmentation Methods

The same-split experiments include U-Net, UNet++, DeepLabV3+, and local MarsLS-Net-style, DualSwin-style lite, and SegFormer-lite references. The style/lite models are architectural approximations rather than official author-released implementations. DeepLabV3+ obtains the highest single-seed and three-seed mean overlap scores, so the present results do not support a claim that TRB-Net is the best-performing model.

TRB-Net instead tests an explicit terrain-residual pathway and boundary output. All conclusions are restricted to the local MMLSv2 split. The official geographically isolated set was not available in the local package used for these experiments, and external-region robustness remains untested.

8.3. Error Sources and Uncertainty Diagnosis

Visual inspection suggests recurring false positives near crater rims, canyon walls, rocky outcrops, and steep non-landslide slopes, and recurring false negatives in small, low-contrast, or fragmented mask regions. Because these geomorphic error categories were not annotated across all tiles, they are reported as observations rather than frequencies or causal findings. The quantitative terrain-only result independently confirms extensive foreground overprediction when RGB is absent.

Predictive entropy is concentrated around many visible mask transitions, and mean entropy correlates with pixel error rate ($r = 0.69$). This supports entropy as a tile-prioritization diagnostic on the local test split, not as a complete calibration assessment. Expected calibra-

tion error, Brier score, ensembles, or cross-region tests would be needed before interpreting the probabilities as calibrated uncertainty.

8.4. Influence of Landslide Area and Boundary-to-Area Ratio

Per-tile foreground IoU has a broader distribution for tiles with a small ground-truth foreground ratio than for large-foreground tiles. This is an observed association in the local split. A geometric explanation is plausible because boundary pixels constitute a larger fraction of a small object, but object-level boundary-to-area ratios were not separately annotated to test that explanation causally.

The result shows why global metrics should be accompanied by per-tile distributions and boundary diagnostics. These analyses reveal heterogeneous tile behavior but do not establish that one architecture is universally preferable for small or large landslides.

8.5. Implications for Remote-Sensing Landslide Mapping

The results support a narrow methodological implication: multimodal Martian segmentation should report both aggregate overlap and diagnostics tied to channel groups, tile heterogeneity, boundaries, and uncertainty. The explicit RGB–terrain separation makes channel-removal and fusion ablations easier to define than undifferentiated early fusion.

They do not yet support operational large-area mapping. A stronger benchmark would pair the present diagnostics with geographically isolated evaluation, independent regions, calibration measures, and expert review of object-level errors.

8.6. Computational Cost and Large-Scale Mapping Considerations

TRB-Net-Compact contains 5.255 million parameters. It is smaller than the local U-Net and UNet++ implementations but larger than DeepLabV3+ and the three style/lite baselines in Table 4; it is therefore not the most parameter-efficient model tested. Parameter count is reported as the reproducible complexity indicator because standardized inference-time and memory benchmarks were not performed.

Any future large-area application would additionally require co-registration, tiling, and post-processing of RGB, DEM, slope, thermal inertia, and grayscale products. These untested data-engineering and runtime requirements are why the manuscript does not claim large-scale mapping readiness.

8.7. Practical Reproducibility Considerations

The manuscript reports the local split, separates compact and multiscale configurations, selects thresholds only on validation data, and identifies style/lite baselines as local approximations. The three-seed comparison reduces dependence on one initialization but does not substitute for multiple spatial splits or isolated-region testing.

Implementation scripts, configurations, and trained weights are available from the corresponding author upon reasonable request. A public release should additionally remove local paths and include preprocessing, normalization, training, evaluation, and figure-generation commands.

9. Limitations and Future Work

Several limitations should be considered when interpreting the results. First, the experiments use the locally available MMLSv2 train/validation/test split. The published dataset also contains a geographically isolated set, but that subset was not present as a separate folder in the local package used for these experiments. This paper therefore does not claim external-region generalization or complete exclusion of regional spatial leakage. Evaluation on the official isolated set or independently prepared Martian regions is required before making such claims.

Second, the conclusions depend on the MMLSv2 data format, annotation style, channel preprocessing, and local sample distribution. Applying the framework to other Martian regions may introduce domain shifts caused by different illumination conditions, dust cover, crater density, surface roughness, sensor products, DEM quality, and co-registration accuracy. The relative value of RGB, DEM, slope, thermal inertia, and grayscale channels may also change across geomorphic settings. Therefore, transfer to other datasets should include local validation, uncertainty analysis, and, where possible, expert geomorphological checking rather than direct deployment of the present checkpoint.

Third, the MMLSv2 tiles have a fixed size of 128×128 pixels. This tile size is suitable for controlled benchmark training, but it limits the amount of large-scale geomorphic context available to the network. Some landslides extend across broader terrain units, and their interpretation may depend on regional slope geometry, runout direction, source area, and depositional setting. Future work could combine tile-level segmentation with larger contextual windows or multi-scale inference so that local texture and regional geomorphic structure are considered jointly.

Fourth, TRB-Net is not the numerically best model in the same-split comparison. DeepLabV3+ achieves the highest mean overlap score. The contribution of TRB-Net is the explicit terrain-residual and boundary-assisted design and its diagnostic evaluation, not a universal replacement for generic segmentation architectures or proof of improved human interpretability.

Finally, the present uncertainty analysis is diagnostic rather than fully probabilistic. Predictive entropy is useful for visualizing ambiguous areas and identifying difficult tiles, but it does not by itself guarantee calibrated probabilities. Future work should evaluate calibration, test-time augmentation uncertainty, ensemble uncertainty, and cross-region uncertainty behavior. Additional field-inspired geomorphic descriptors, such as slope position, local relief, or runout-context features, may also improve the connection between deep segmentation outputs and planetary landslide interpretation.

10. Conclusions

This study evaluates TRB-Net, a terrain-residual and boundary-assisted multimodal network, on a local MMLSv2 split. The architecture separates RGB and auxiliary terrain inputs. It injects transformed terrain features into RGB-dominant representations through residual fusion and includes an auxiliary boundary output. These operations encode explicit modeling assumptions; they do not establish a geological causal mechanism.

On the local MMLSv2 test split, TRB-Net-Compact achieves an mIoU of 0.8060, foreground IoU of 0.7502, F1-score of 0.8573, precision of 0.8473, and recall of 0.8676 at threshold 0.55. In the separate single-run multiscale ablation series, the reference configuration scores higher than the tested RGB-only, terrain-only, alternative-fusion, no-ASPP, and no-boundary variants on the principal overlap metrics; the no-DEM variant is the stated exception.

DeepLabV3+ obtains the highest same-split overlap scores. TRB-Net-Compact contains 5.255 million parameters and provides an explicit terrain and boundary pathway for diagnostic analysis. The conclusions are limited to the local split; cross-region generalization, calibrated interpretability, and large-area mapping performance remain to be tested.

Author Contributions: Conceptualization, Y.L. and J.H.; methodology, Y.L.; software, Y.L.; validation, Y.L. and H.Z.; formal analysis, Y.L. and Y.Y.; investigation, Y.L., Y.W. and Y.Z.; resources, J.H.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, J.H., Y.W., Y.Z., Y.Y. and H.Z.; visualization, Y.L.; supervision, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Key Research and Development Program of China (grant numbers 2022YFF0503102 and 2022YFF0503100).

Data Availability Statement: The MMLSv2 dataset used in this study is publicly available from the repository associated with Paheding et al. [6]. The experimental split used in this work contains 465 training samples, 66 validation samples, and 133 local test samples. The geographically isolated test set described by the public MMLSv2 documentation was not available as a separate local folder in the package used for these experiments, and isolated-test results are therefore not claimed in this study. The MGS MOLA MEGDR topography used to generate Figure 1 is publicly available from the PDS/USGS archive. Source data for the quantitative comparison figures are provided with the manuscript figure files. The implementation scripts, model configuration files, and trained weights used in this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the providers of the MMLSv2 dataset and the PDS/USGS MOLA archive for making the remote-sensing and topographic data publicly available.

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

References

1. Lucchitta, B.K. Landslides in Valles Marineris, Mars. *J. Geophys. Res. Solid Earth* **1979**, *84*, 8097–8113.
2. Quantin, C.; Allemand, P.; Delacourt, C. Morphology and geometry of Valles Marineris landslides. *Planet. Space Sci.* **2004**, *52*, 1011–1022.
3. Brunetti, M.T.; Guzzetti, F.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Mancinelli, P.; Komatsu, G.; Borselli, L. Analysis of a new geomorphological inventory of landslides in Valles Marineris, Mars. *Earth Planet. Sci. Lett.* **2014**, *405*, 156–168.
4. Mazzanti, P.; De Blasio, F.V.; Di Bastiano, C.; Bozzano, F. Inferring the high velocity of landslides in Valles Marineris on Mars from morphological analysis. *Earth Plan. Space* **2016**, *68*, 195.
5. Paheding, S.; Reyes-Angulo, A.; Rajaneesh, A.; Sajinkumar, K.S.; Oommen, T. MarsLS-Net: Martian Landslides Segmentation Network and Benchmark Dataset. In Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2024; pp. 8236–8245.
6. Paheding, S.; Reyes-Angulo, A.; Ramos, L.T.; Sappa, A.D.; Rajaneesh, A.; Hiral, P.B.; Kumar, K.S.S.; Oommen, T. MMLSv2: A Multimodal Dataset for Martian Landslide Detection in Remote Sensing Imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Denver, CO, USA, 3–7 June 2026; pp. 10329–10338.
7. Kabir, S.; Ehsan, A.M.A.; Rifi, I.A.; Reza, M.K. DualSwinFusionSeg: Multimodal Martian Landslide Segmentation via Dual Swin Transformer with Multi-Scale Fusion and UNet++. *arXiv* **2026**, arXiv:2603.14132.
8. Crosta, G.B.; De Blasio, F.V.; Frattini, P. Global scale analysis of Martian landslide mobility and paleoenvironmental clues. *J. Geophys. Res. Plan.* **2018**, *123*, 872–891.
9. Malin, M.C.; Bell, J.F.; Cantor, B.A.; Caplinger, M.A.; Calvin, W.M.; Clancy, R.T.; Edgett, K.S.; Edwards, L.; Haberle, R.M.; James, P.B.; et al. Context Camera Investigation on board the Mars Reconnaissance Orbiter. *J. Geophys. Res. Plan.* **2007**, *112*, E05S04.
10. Smith, D.E.; Zuber, M.T.; Frey, H.V.; Garvin, J.B.; Head, J.W.; Muhleman, D.O.; Pettengill, G.H.; Phillips, R.J.; Solomon, S.C.; Zwally, H.J.; et al. Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *J. Geophys. Res. Plan.* **2001**, *106*, 23689–23722.
11. Christensen, P.R.; Jakosky, B.M.; Kieffer, H.H.; Malin, M.C.; McSween, H.Y., Jr.; Nealson, K.; Mehall, G.L.; Silverman, S.H.; Ferry, S.; Caplinger, M.; et al. The Thermal Emission Imaging System (THEMIS) for the Mars 2001 Odyssey mission. *Space Sci. Rev.* **2004**, *110*, 85–130.
12. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.
13. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *International Workshop on Deep Learning in Medical Image Analysis*; Springer: Cham, Switzerland, 2018; pp. 3–11.
14. 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, Munich, Germany, 8–14 September 2018; pp. 801–818.
15. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
16. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 11–17 October 2021; pp. 10012–10022.

17. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin Transformer V2: Scaling Up Capacity and Resolution. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 12009–12019.
18. Ghorbanzadeh, O.; Xu, Y.; Ghamisi, P.; Kopp, M.; Kreil, D. Landslide4Sense: Reference Benchmark Data and Deep Learning Models for Landslide Detection. *arXiv* **2022**, arXiv:2206.00515.
19. Zhao, C.; Li, L.; Wang, Y.; Li, X.; Xu, C.; Song, Y.; Ren, D.; Xiao, C. D2FLS-Net: Dual-Stage DEM-Guided Fusion Transformer for Landslide Segmentation. *PLoS ONE* **2025**, *20*, e0337412.
20. Jiang, P.; Ma, Z.; Mei, G. Review article: Deep learning for potential landslide identification: Data, models, applications, challenges, and opportunities. *Nat. Hazards Earth Syst. Sci.* **2026**, *26*, 487–529.
21. Qin, D.; Li, Q.; Fang, L. Fanet: Landslide recognition in remote sensing images based on multi-source data. *Environ. Earth Sci.* **2026**, *85*, 205.
22. Du, B.; Huang, W.; Ye, J.; Tong, B.; Yin, Y. A Deformable Dual-Branch Visual State-Space Network for Landslide Identification with Multi-Scale Recognition and Irregular Boundary Enhancement. *Remote Sens.* **2026**, *18*, 707.
23. Rajaneesh, A.; Vishnu, C.L.; Oommen, T.; Rajesh, V.J.; Sajinkumar, K.S. Machine learning as a tool to classify extra-terrestrial landslides: A dossier from Valles Marineris, Mars. *Icarus* **2022**, *376*, 114886.
24. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
25. 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 2018), Munich, Germany, 8–14 September 2018; pp. 3–19.
26. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
27. Milletari, F.; Navab, N.; Ahmadi, S.-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
28. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
29. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the 7th International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
30. Berman, M.; Rannen Triki, A.; Blaschko, M.B. The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-over-Union Measure in Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4413–4421.

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