frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 16 October 2025
DOI 10.3389/feart.2025.1685685

OPEN ACCESS

EDITED BY
Sheng Nie,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Qinjun Wang,
Chinese Academy of Sciences, China
Linhai Jing,
China University of Geosciences, China

*CORRESPONDENCE
Suhong Liu,
liush@bnu.edu.cn

RECEIVED 14 August 2025
ACCEPTED 25 September 2025
PUBLISHED 16 October 2025

CITATION
Zhou H, Liu S, Zhou C and Ma Z (2025) Scale-adaptive and mask refinement modules for accurate alluvial fan boundary detection in remote sensing data.
Front. Earth Sci. 13:1685685.
doi: 10.3389/feart.2025.1685685

Scale-adaptive and mask refinement modules for accurate alluvial fan boundary detection in remote sensing data

Hongying Zhou[1,2], Suhong Liu[1]*, Chengkai Zhou[3] and Zhiguo Ma[2]

[1]Department of Geomorphology and Remote Sensing, Faculty of Geographical Science, Beijing Normal University, Beijing, China, [2]Research Institute of Petroleum Exploration and Development, PetroChina, Beijing, China, [3]Power China Urban Planning and Design Institute Co., Ltd., Guangzhou, China

Introduction: Alluvial fans are crucial geomorphic features in arid regions, playing key roles in geomorphic evolution, hydrological modeling, and land-use planning. However, their irregular morphology and multi-scale characteristics make accurate boundary delineation challenging for conventional remote sensing methods.
Methods: To overcome these limitations, this study proposes a multi-module enhanced Mask R-CNN framework that integrates topographic and spectral information for precise alluvial fan recognition. The model consists of a Topographic–Spectral Fusion (TSF) module, a Scale-Adaptive Module (SAM), and a Mask–Boundary Refinement (MBR) module, jointly designed to improve recognition accuracy and structural detail preservation.
Results: Experiments based on multi-source remote sensing imagery and terrain data show that the proposed model achieves an accuracy of 91.7%, precision of 89.8%, recall of 88.5%, and F1-score of 89.1% in full-region classification. For segmentation, the model attains a mean intersection over union (mIoU) of 81.5% and a boundary F1-score of 80.4%. Ablation experiments confirm that the TSF module enhances spatial–structural modeling, while the MBR module improves boundary fitting.
Discussion: The results demonstrate that the proposed framework provides robust and transferable performance across different fan size categories, achieving a minimum false negative rate of 3.9%. The method offers both theoretical value and practical applicability for accurate alluvial fan recognition in arid regions.

KEYWORDS

computer vision, alluvial fans segmentation, multi-scale feature extraction, boundary-aware mask refinement, high-resolutionimage analysis

1 Introduction

Alluvial fans are fan-shaped depositional systems formed by rapid sedimentation at mountain outlets due to abrupt decreases in hydrodynamic energy. These fans are widely distributed in arid and semi-arid regions and hold significant implications in

sedimentology, geomorphology, and resource geology (Ghahraman, 2024). In fields such as petroleum exploration, groundwater development, debris flow monitoring, and the study of modern depositional environments, the accurate identification and delineation of alluvial fans—representing major coarse-grained depositional units—play a crucial role in reservoir prediction, watershed modeling, and disaster prevention planning (Shoshta and Marh, 2023). Therefore, precise and efficient recognition of the spatial distribution and boundary features of alluvial fans is of both theoretical and practical importance for enhancing resource detection and geomorphological understanding. Traditional methods for identifying alluvial fans have predominantly relied on field geological surveys, manual interpretation of remote sensing imagery, and spectral index techniques. Although feasible for small-scale investigations, these approaches suffer from substantial limitations, including insufficient spatial coverage, heavy reliance on expert experience, high subjectivity, and poor automation and batch-processing capabilities (Ghahraman and Nagy, 2023). Particularly in large-scale complex geomorphic regions, such methods fail to meet the demands for efficient, objective, and fine-grained recognition, thereby hindering the broader application and scalability of alluvial fan studies.

In recent years, deep learning techniques have been widely applied in remote sensing image analysis, with architectures such as convolutional neural networks (CNNs), U-Net, and MASK R-CNN achieving notable success in urban boundary extraction and disaster detection tasks (Lin et al., 2022). However, several technical bottlenecks remain when these models are applied to the identification of alluvial fans (Lv et al., 2023). On one hand, most existing models rely solely on optical imagery, limiting their ability to capture three-dimensional geomorphic structures and neglecting critical topographic features such as slope and aspect. On the other hand, current network architectures are typically designed for general object detection tasks and lack structural adaptation and boundary refinement mechanisms tailored to sedimentary fans. As a result, their accuracy and generalizability are constrained when dealing with alluvial fans of varying stages and source materials. To address the challenges of low automation, weak spatial structural representation, and inadequate utilization of topographic information in current alluvial fan recognition methods, an intelligent recognition framework integrating spectral and topographic information is proposed in this study. Specifically, a six-channel remote sensing input is constructed by combining RGB bands with terrain bands, including digital elevation model (DEM), slope, and aspect, to comprehensively encode both spectral and spatial geometric features of geomorphology. On this basis, the MASK R-CNN deep neural network architecture is modified to enhance the precision and robustness of boundary detection and morphological characterization for sedimentary bodies. The proposed method enables automatic extraction of alluvial fan regions while balancing recognition accuracy and spatial consistency, thereby significantly improving the intelligence level of remote sensing interpretation. The main contributions of this study are summarized as follows:

1. A six-channel remote sensing input mechanism that fuses spectral and topographic information is proposed. For the first time, DEM, slope, and aspect are jointly modeled with RGB bands to enhance the spatial perception of alluvial fan geomorphology;
2. The MASK R-CNN architecture is modified to accommodate six-channel input, including reconstruction of the initial convolution layer, introduction of a slope-guided dynamic anchor configuration mechanism, and development of a boundary-aware mask optimization strategy, thereby improving adaptability to scale variation and boundary complexity;
3. A high-quality labeled remote sensing dataset is constructed, and empirical studies are conducted in representative alluvial fan areas located in the Junggar Basin and the southern margin of the Qilian Mountains, demonstrating the effectiveness and stability of the method under diverse geomorphic conditions.

2 Related work

2.1 Remote sensing-based identification of alluvial fans

Alluvial and debris flow fans, as typical coarse-grained depositional systems developed at the piedmont zones, are widely distributed in arid, semi-arid, and mountainous regions, holding significant implications for geomorphic evolution and resource-environmental applications. With the advancement of remote sensing technologies, increasing efforts have been devoted to the identification and extraction of alluvial fans using multi-source satellite data, leading to the formation of a preliminary technical framework (Zhou et al., 2022). Existing studies have primarily focused on three aspects: delineation of fan boundaries, extraction of morphological parameters, and spatial analysis of depositional evolution stages. The mainstream methodologies can be categorized into three types: visual interpretation, spectral index-based approaches, and morphological parameter analysis. Visual interpretation relies on color, texture, shape, and topographic context within satellite imagery for manual delineation, and has been widely applied in early studies (Miliaresis and Argialas, 2000). However, this method suffers from low efficiency, strong dependence on expert experience, and limited scalability. Spectral index methods utilize the differences in spectral signatures of vegetation, water, and soil to assist in boundary extraction. Gao et al. proposed and validated the use of the normalized difference water index (NDWI), computed from near-infrared and shortwave-infrared bands, to detect surface water and vegetation moisture, providing foundational support for alluvial fan delineation in arid zones (Gao, 1996). Thannoun employed principal component analysis (PCA), band ratioing, and false-color composites based on Landsat-7 ETM + imagery to extract fan boundaries in northern Iraq (Thannoun et al., 2016). These methods offer simplicity and are suitable for preliminary large-area delineation, yet are highly sensitive to imaging conditions and surface cover, and are limited in capturing inherent geomorphic structures.

Morphological parameter analysis utilizes remote sensing imagery or DEMs to extract geometric attributes such as slope, curvature, and spatial extent (Zhang et al., 2022). Thresholding or clustering algorithms are then applied to identify fan morphologies. Babič et al. developed an automated framework

using DEMs to evaluate key parameters representing complex geomorphic characteristics, such as relative positioning within the surrounding terrain, and employed five machine learning algorithms to detect Slovenian torrential fans (Babič et al., 2021). Nevertheless, due to the morphological heterogeneity of alluvial fans formed under diverse provenance and climatic settings, conventional morphological models often lack generalization and robustness. To address these gaps, a deep learning-based method is proposed that integrates spectral and topographic features and supports end-to-end automatic recognition. This approach aims to overcome the shortcomings of traditional techniques by enhancing the modeling of spatial geometric information and enabling intelligent interpretation of complex sedimentary fans, thereby facilitating the paradigm shift from rule-based to data-driven geomorphological mapping.

2.2 Deep learning in remote sensing-based geomorphological recognition

With recent advancements in the spatial resolution and revisit frequency of satellite imagery, deep learning has emerged as a powerful tool for the extraction and intelligent recognition of geomorphic features. CNNs known for their strong capability in spatial feature modeling, have been widely adopted for classification, detection, and segmentation tasks in remote sensing (Mei et al., 2024; Li et al., 2024). Among them, encoder-decoder architectures such as U-Net (Wang and Li, 2024) and the DeepLab series (Wang et al., 2024) have achieved notable performance in semantic segmentation, enabling pixel-level delineation of geomorphic units. U-Net employs skip connections to fuse multi-scale contextual information, making it suitable for detecting fans with clear boundaries and high connectivity. DeepLabv3+ enhances the perception of complex textures and scale variations through atrous convolution and spatial pyramid pooling.

These models have been successfully applied in geological hazard monitoring, land-use classification, and ecological zoning. Compared to semantic segmentation models, Mask R-CNN (Jiang et al., 2024) offers the combined capabilities of object detection and instance segmentation, with superior boundary localization and structural expression. Its applications span urban boundary extraction (Hou and Li, 2024; Ismael and Sadeq, 2025), landslide detection, and debris flow mapping (Wan et al., 2024). Studies have demonstrated that Mask R-CNN's region proposal network (RPN) and mask branch are effective in capturing spatially complex and structurally ambiguous geomorphic entities, making it particularly suitable for targets with prominent spatial boundaries but diverse morphological characteristics. Therefore, a remote sensing recognition framework that integrates multi-source data, demonstrates structural sensitivity, and supports regional generalization is urgently needed. By incorporating topographic parameters and constructing feature extraction mechanisms specific to fan identification, and by combining instance segmentation with boundary refinement strategies, model capabilities in delineating complex sedimentary fans can be significantly enhanced. This study introduces a modified Mask R-CNN model tailored for alluvial fan recognition, aiming to achieve high-accuracy and robust performance in geomorphic identification tasks.

2.3 Multi-source fusion strategies for remote sensing data

As a core direction in applied remote sensing, geomorphological recognition has gradually shifted from reliance on single-source data to multi-source information integration. Traditional remote sensing analyses primarily leverage spectral features from visible and near-infrared bands to infer surface materials and fan structures. However, for complex geomorphic types such as sedimentary bodies, fluvial networks, and desert fans, spectral information alone is insufficient for accurate characterization and robust identification. To address this, increasing attention has been paid to integrating topographic variables—such as DEM, slope, and aspect—with multispectral imagery, enabling more comprehensive modeling of fan morphology, slope characteristics, and structural evolution (Li et al., 2020). Current mainstream fusion strategies can be categorized into three types: multi-source stacked input, multi-channel encoding, and deep feature fusion. Among them, directly stacking spectral and topographic variables into multi-channel input images has become the most widely adopted approach in deep learning models. This strategy preserves the original spatial resolution and positional alignment of each data source, simplifies preprocessing, and provides high-dimensional and complementary discriminative features to neural networks (Lyu et al., 2021).

Multi-channel inputs significantly enhance model sensitivity to geometric and morphological features, improve discrimination in complex backgrounds, and boost generalization performance—particularly beneficial in tasks involving blurred fan boundaries and large scale variations. For instance, six-band composite inputs have demonstrated advantages in various geomorphic scenarios. In aeolian desert fan recognition, DEM and slope information help delineate dune orientations and morphologies (Udin et al., 2019); in fluvial and alluvial plain analysis, aspect and elevation gradients are crucial for identifying floodplain boundaries and channel distributions (Odunuga and Raji, 2018); and in alluvial fan recognition, slope gradients and radial dispersion patterns derived from DEM can clearly distinguish fan structures. Studies have shown that such spectral-topographic joint input strategies significantly improve sensitivity to critical features, such as spatial boundaries, fan-edge gullies, and slope discontinuities, thereby enhancing segmentation accuracy and boundary delineation.

To overcome these challenges, a fusion modeling framework based on six-channel remote sensing inputs is proposed. RGB spectral bands and three topographic variables—DEM, slope, and aspect—are systematically integrated. The network input layer is reconstructed to accommodate the high-dimensional input. The theoretical foundation of this strategy lies in the complementarity between spectral and spatial geometric information. By enabling both data-level and structure-level fusion, the model's representation of complex depositional boundaries and morphologies is enhanced, providing a structured solution for intelligent recognition of alluvial and other sedimentary fans.

TABLE 1 Statistics of remote sensing data and annotated samples.

Data type	Junggar basin	Southern qilian	Total
LANDSAT-7 RGB imagery (2018–2022)	24 scenes	18 scenes	42 scenes
GDEMV2 DEM and derived layers (2018–2022)	24 tiles	18 tiles	42 tiles
Six-channel image tiles (Training set)	1120	733	1853
Six-channel image tiles (Validation set)	122	84	206
Six-channel image tiles (Test set)	115	85	200

3 Materials and methods

3.1 Data collection

The remote sensing data employed in this study were primarily derived from LANDSAT-7 ETM + optical imagery and GDEMV2 topographic datasets, covering two typical arid-region alluvial fan development zones located in the northwestern margin of the Junggar Basin and the southern margin of the Qilian Mountains, as shown in Table 1. The optical imagery was acquired from the United States Geological Survey (USGS) Earth Explorer platform, with acquisition dates ranging from 2018 to 2022. Priority was given to scenes captured between June and October with cloud coverage less than 5%, thereby ensuring clear surface observations free from cloud contamination. The selected LANDSAT-7 ETM + images included red (R), green (G), and blue (B) bands with a spatial resolution of 30 m, offering robust fan representation and consistent large-scale image coverage. After downloading, all scenes underwent radiometric calibration, atmospheric correction, and geometric registration to ensure spectral consistency and spatial alignment across different years and regions. The topographic data were collected over a similar time span and were uniformly sourced from the GDEMV2 dataset released by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. This dataset also provides a spatial resolution of 30 m and was constructed through the fusion of ASTER GDEM, SRTM, and domestic photogrammetric survey data, ensuring high elevation accuracy and regional consistency. In this study, the original DEM layers were used as the base, from which slope and aspect layers were derived using raster-based differential algorithms. These layers collectively formed the topographic channels. To enhance the representation of three-dimensional geomorphological features, all topographic data were resampled and aligned at the pixel level, and reprojected to the WGS 84 UTM coordinate system. Furthermore, the elevation and its derivatives were normalized to the $[0, 1]$ interval using z-score normalization to prevent gradient imbalance during network training due to extreme terrain variation.

During the data fusion stage, the RGB optical bands and the topographic channels (DEM, slope, aspect) were concatenated along the channel dimension to construct a six-channel remote sensing input. The resulting composite imagery preserved original spectral and textural information while introducing structural priors, thereby enhancing the model's capacity to perceive alluvial fan morphologies and boundary features. To build a high-quality labeled dataset, a manual annotation process was conducted using the ArcGIS platform, based on LANDSAT imagery and DEM data collected between 2019 and 2021. Through expert interpretation of visual spectral features and topographic cross-sections, the boundaries of alluvial fans were accurately delineated. The labeled regions spanned multiple representative fans characterized by different sediment sources, depositional phases, and distinct geomorphic configurations of fan structures. The composite images and their corresponding vector labels were then segmented into multiple image tiles at various scales, with resolutions ranging from 512×512 to 2048×2048 pixels. This approach enabled the deep learning model to simultaneously capture local detail and global structure. The final dataset comprised 1853 six-channel tiles for training, 206 for validation, and 200 for testing, effectively covering diverse geomorphic patterns, background noise, and depositional scenarios, and providing a solid foundation for model training and generalization.

3.2 Data preprocessing and augmentation strategy

To construct a high-quality training dataset suitable for multi-source remote sensing inputs, systematic preprocessing and augmentation procedures were applied to both the raw optical and topographic data prior to model training. To enrich the input beyond traditional RGB imagery, two spatial derivatives—slope and aspect—were extracted from the DEM to supplement the missing geometric information. Slope quantifies the steepness of elevation changes and was computed using the central difference method as follows:

$$\text{Slope}(x, y) = \tan^{-1}\left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\right),$$

where $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ represent elevation gradients along the horizontal and vertical directions, approximated using a 3×3 Sobel operator. Aspect describes the main orientation of the slope and was calculated as:

$$\text{Aspect}(x, y) = \tan^{-1}\left(\frac{\frac{\partial z}{\partial y}}{\frac{\partial z}{\partial x}}\right).$$

To avoid discontinuities caused by the circular nature of directional angles within the 360° domain, the aspect was projected

onto the unit circle and normalized as:

$$\text{Aspect}_{\text{norm}} = \frac{1}{2\pi} \cdot (\text{Aspect mod } 2\pi).$$

This transformation ensures consistency in scale and periodic stability across all input channels, facilitating the network's ability to learn directional patterns of surface slopes. To eliminate inter-band scale discrepancies, each of the six input channels was normalized individually. For the RGB optical bands, standard score normalization was applied:

$$x' = \frac{x-\mu}{\sigma},$$

where μ and σ denote the mean and standard deviation of the respective channel. For topographic bands, a terrain-guided normalization strategy was adopted to suppress extreme values in high-relief regions. Slope values exceeding a threshold τ (e.g., 30°) were penalized using a suppression term:

$$x' = \frac{x}{1+\alpha \cdot \max(0, x-\tau)},$$

where α is a scaling factor introduced to mitigate gradient instability caused by abrupt slope variations. To improve model generalization to the morphological diversity of alluvial fans, a series of data augmentation strategies were designed and applied dynamically during training. Let $I(x,y)$ denote the original image and $I^*(x,y)$ the augmented version. First, to simulate directional variability of fans within terrain, affine rotation was performed:

$$I^*(x,y) = I\left(R_\theta \cdot [x,y]^T\right), \quad \theta \sim \mathcal{U}(-30°, 30°),$$

where R_θ is the 2×2 rotation matrix and θ is sampled from a uniform distribution over $[-30°, 30°]$. Next, scale transformation was introduced to account for spatial size variation of fans:

$$I^*(x,y) = I(s \cdot x, s \cdot y), \quad s \sim \mathcal{U}(0.75, 1.25),$$

where s is a scaling factor sampled from $\mathcal{U}(0.75, 1.25)$, and interpolation was used to maintain the original image size. To model illumination variability in remote sensing imagery, brightness perturbation was applied:

$$I^*(x,y) = I(x,y) \cdot (1+\delta), \quad \delta \sim \mathcal{U}(-0.15, 0.15),$$

where δ is a perturbation factor controlling global brightness variation. Additionally, random image flipping was used to augment directional symmetry:

$$I^*(x,y) = I(W-x, y)$$

for horizontal flipping, and

$$I^*(x,y) = I(x, H-y)$$

For vertical flipping, where W and H denote the image width and height, respectively. All augmentation strategies were applied with randomized combinations during each training epoch, effectively expanding the training distribution and improving model robustness and recognition accuracy under complex geomorphic conditions. After applying these augmentation strategies, the number of training samples was increased from 1,853 original six-channel image tiles to approximately 5,559 augmented samples, providing a threefold expansion of the dataset and ensuring greater diversity for model learning.

3.3 Proposed method

3.3.1 Overall

A multi-module recognition framework integrating spectral information and topographic structure was constructed in this study, aiming to achieve automatic recognition and fine boundary segmentation of alluvial fans. As shown in Figure 1, the model receives as input a six-channel composite image consisting of conventional RGB optical bands and three terrain-derived channels (elevation, slope, and aspect) from DEM data. The input is first passed to the topographic-spectral fusion input module, where the six-channel image is encoded through a modified ResNet101 backbone, enabling unified extraction of low-level features. This stage is critical for simultaneously capturing surface texture and spatial geometric features, allowing the model to perceive the gradient patterns and radial dispersal structure characteristic of alluvial fans. The extracted features are then fed into the scale-adaptive optimization module, which contains a terrain-gradient-driven anchor generation mechanism that adaptively proposes candidate regions according to local slope variations. Additionally, the FPN structure and geomorphic feature enhancement links are introduced to improve the model's response to fan boundaries and small-scale lobes. The proposed candidate regions are subsequently forwarded to the mask prediction branch, entering the mask refinement and boundary-aware segmentation module. At this stage, boundary attention mechanisms are introduced for explicit modeling of fan boundaries, the output mask resolution is increased, and edge-guided loss terms are incorporated to refine boundary fitting and enhance the model's ability to reconstruct complex fan morphologies. The entire workflow forms a structural loop of "feature encoding–scale localization–boundary optimization," where each module collaborates via shared spatial features and gradient feedback. Compared to the traditional three-channel MASK R-CNN, the proposed method significantly improves the model's capacity to model spatial structures of alluvial fans and enhances its accuracy and boundary interpretability under multi-scale and irregular conditions.

3.3.2 Topographic-spectral fusion input module

As shown in Figure 2, the proposed topographic-spectral fusion input module was developed to address the limitations of conventional remote sensing methods that rely solely on RGB imagery, which lack the ability to capture geometric structures. By incorporating multi-source terrain information, this module enables accurate modeling of the complex spatial features of alluvial fans. Structurally, optical imagery and terrain data are integrated at the channel level, resulting in a six-channel input composed of red, green, and blue optical bands, as well as DEM, slope, and aspect channels. To accommodate this high-dimensional input, the initial convolutional layer of the ResNet101 backbone was modified, expanding the original kernel from $7\times7\times3$ to $7\times7\times6$, thereby increasing the input channels from 3 to 6. The output feature map size remains 112×112, with batch normalization and ReLU activation preserved, and residual connections maintained to ensure training stability. In data processing, the terrain channels are normalized. Slope is computed using the central difference method, while aspect is derived via gradient ratio and mapped to the unit circle to maintain continuity and differentiability

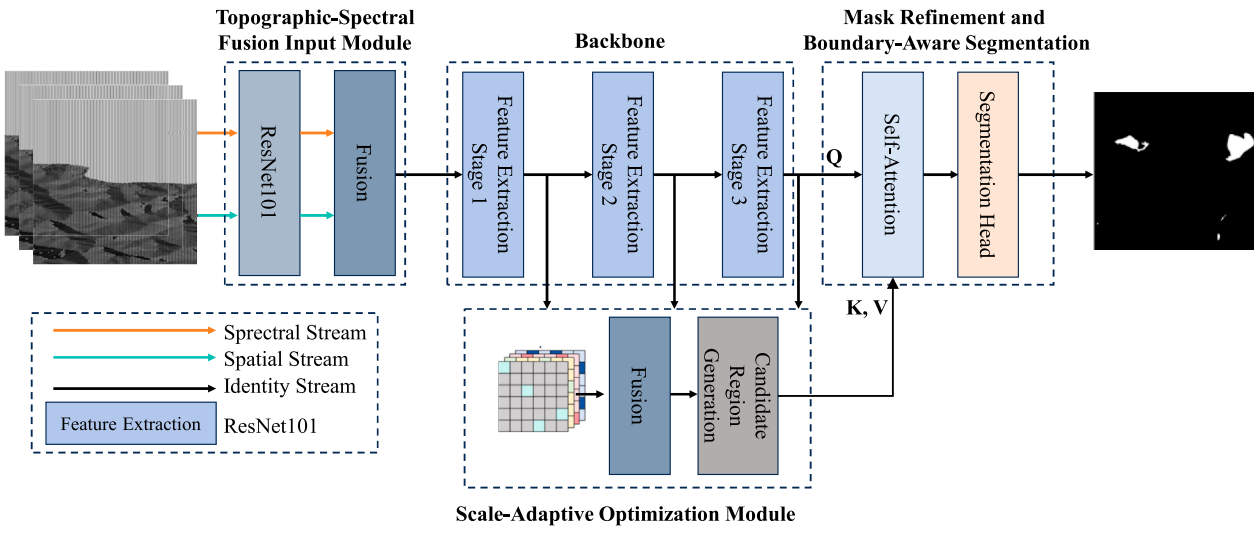

FIGURE 1
The illustrated overall architecture presents a remote sensing information processing framework.

in periodic angular features. Mathematically, for each pixel (x,y), the slope is defined as $\text{Slope}(x,y) = \tan^{-1}\left(\sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2}\right)$, and the aspect is defined as $\text{Aspect}(x,y) = \tan^{-1}\left(\frac{\frac{\partial z}{\partial y}}{\frac{\partial z}{\partial x}}\right)$, with normalization applied as $\text{Aspect}_{\text{norm}} = \frac{1}{2\pi} \cdot (\text{Aspect} \bmod 2\pi)$. These encoded features enable the input image to retain both spectral texture and 3D structural information, significantly enhancing the model's initial geomorphic perception capabilities. Furthermore, a dual-branch feature extraction structure is employed after the backbone, consisting of a spectral stream and a spatial stream. The spectral stream employs a lightweight convolutional encoder and multi-head self-attention to capture inter-band relationships, while the spatial stream adopts Mamba blocks to implement non-local spatial modeling for terrain features. These features are subsequently fused in the cross-dimensional feature enhancement module, where bidirectional attention mechanisms (spectral-to-spatial and spatial-to-spectral) reweight the features. Through the construction of selected spatial queries, terrain priors are used to filter spatial regions highly correlated with fan morphology, which are then utilized to guide the downstream mask prediction process.

From a mathematical perspective, the topographic-spectral fusion module not only extends the input dimensionality but also enables cooperative feature modeling in the latent space. The spectral and topographic channels exhibit complementarity across multiple scales. The terrain-guided attention mechanism can be interpreted as spatial modulation of attention weights $A_{i,j}$, and the optimization objective can be considered as minimizing reconstruction error under spatial-spectral discrepancy. This design significantly improves the model's ability to express critical structural features such as slope inflection, fan dispersion direction, and edge discontinuities, making it particularly suitable for identifying alluvial fan fans with strong geometric organization. Experimental results demonstrate that this module, as the front-end of the feature extraction pipeline, offers superior accuracy and boundary completeness compared to traditional RGB-only input models.

3.3.3 Scale-adaptive optimization module

As shown in Figure 3, the proposed scale-adaptive optimization module was designed to enhance the model's perception of multi-scale morphology and complex boundary structures of alluvial fans. The core idea involves dynamic anchor generation, feature pyramid construction, and channel-wise recalibration guided by local geometric structures, enabling accurate localization of geomorphic units across varying spatial scales, particularly in fan bodies characterized by multi-phase stacking (Zhang et al., 2021). Structurally, the module extends the RPN by incorporating terrain gradient information to reconstruct anchors and employs a multi-scale grouped convolution pathway for scale-adaptive modeling. The input consists of multi-scale feature maps from the third, fourth, and fifth layers of the ResNet101 backbone, denoted as $C_3 \in \mathbb{R}^{256\times128\times128}$, $C_4 \in \mathbb{R}^{512\times64\times64}$, and $C_5 \in \mathbb{R}^{1024\times32\times32}$, respectively. Based on these, a feature pyramid network (FPN) is constructed, with each layer aligned to a unified channel dimension $C = 256$ using 1×1 convolution. Each level of the pyramid is then passed to a geomorphic scale-adaptive sub-module, whose structure is illustrated in the figure. This sub-module contains three parallel branches (Group1, Group2, Group3), each employing combinations of convolution kernels with different receptive fields (e.g., 3×3 and 1×1), where each branch comprises $L = 2$ convolutional layers with output channels set to $C_l = 64$. Outputs are aggregated via average pooling to form scalar guidance coefficients. The core scale selection strategy is governed by the following dynamic reconstruction (Equation 1): given the local gradient $G(x,y)$ in a feature map region, the anchor scale s is defined as

$$s(x,y) = s_0 \cdot \left(1 + \beta \cdot \frac{\partial G(x,y)}{\partial n}\right), \tag{1}$$

where s_0 denotes the base anchor scale, β is a sensitivity factor, and $\frac{\partial G(x,y)}{\partial n}$ represents the rate of local variation along the gradient

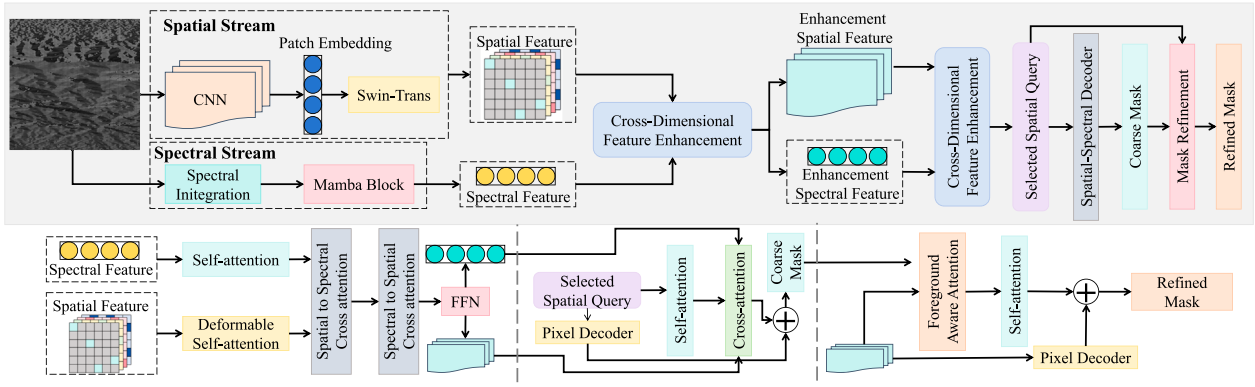

FIGURE 2
The figure illustrates the architecture of the topographic-spectral fusion input module. Spatial and spectral features are extracted via parallel spatial and spectral streams, respectively, with terrain attributes (e.g., elevation, slope) and remote sensing spectral bands (e.g., texture, reflectance). These features are integrated using the cross-dimensional feature enhancement module and fed into a spatial-spectral decoder equipped with self-attention and deformable attention mechanisms.

direction. This formulation theoretically ensures a monotonic response between anchor scale and terrain gradient, generating denser anchors in areas of slope discontinuity while suppressing redundant candidates in flat regions. Furthermore, a channel-wise recalibration (Equation 2) is introduced via a geomorphology-aware weighting map $M \in \mathbb{R}^{C\times1\times1}$, used to enhance feature channels through the Hadamard product, defined as

$$\hat{F} = F \odot \sigma(\mathrm{ReLU}(W_2 \cdot \mathrm{AvgPool}(W_1 \cdot F))), \tag{2}$$

where F denotes the original feature map, W_1 and W_2 are learnable parameters, and σ is the sigmoid function. This non-linear channel mapping allows feature-level emphasis and region-level scale redistribution, enabling selective amplification of geomorphic responses. Combined with the topographic-spectral fusion module, this component constitutes a complete spatial structure modeling path. The former provides rich geometric priors, while the latter adaptively adjusts the perceptual scale at the regional level. Theoretically, this can be interpreted as implicit modeling of a cross-scale attention (Equation 3), with an equivalent objective function defined as

$$\min_{s,M} \mathbb{E}_{(x,y)\sim\Omega} \left| \mathcal{F}_{\mathrm{fusion}}(x,y;M) - \mathcal{F}_{\mathrm{anchor}}(x,y;s) \right|_2^2, \tag{3}$$

where $\mathcal{F}_{\mathrm{fusion}}$ denotes the fused spatial-channel attention features, $\mathcal{F}_{\mathrm{anchor}}$ denotes the anchor-based region response, and Ω represents the training sample set. In the task of alluvial fan recognition, this multi-scale modeling mechanism accommodates morphological differences in fan width, length, and dispersion direction across provenance types, improving robustness and boundary fitting precision. Experimental results also demonstrated strong adaptability and stability across different study areas, with higher recall and accuracy observed in complex fan geometries.

3.3.4 Mask refinement and boundary-aware segmentation

The mask refinement and boundary-aware segmentation module was designed to address limitations in traditional MASK R-CNN architectures, particularly the issues of boundary blurring, mask aliasing, and inadequate segmentation precision in geomorphic recognition tasks. Structural enhancements were introduced to accommodate the highly variable outlines and complex edge curvatures of alluvial fans. This module is built upon the original mask branch of MASK R-CNN and embeds a learnable boundary attention channel along with a hierarchical mask refinement mechanism. Additionally, a support-query consistency matching mechanism was introduced to achieve high-fidelity reconstruction of fan boundary structures.

As shown in Figure 4, the module receives high-resolution, multi-scale feature maps from the scale-adaptive optimization module, denoted as $F \in \mathbb{R}^{C\times H\times W}$, where $C = 256$ and $H = W = 28$. To enhance mask expressiveness, the feature maps are upsampled to 56×56 via a pixel decoder structure comprising two branches: a primary path for mask prediction and an auxiliary path for boundary-aware guidance. These are fused via residual connections. In the primary path, two standard transformer blocks—each with $C = 128$ channels and 8 attention heads—are employed to reconstruct mask features using multi-head self-attention and feed-forward networks. Positional encoding consistency is maintained to enhance contextual awareness in the mask representation. In the boundary-guided path, a pseudo-boundary map $E(x,y)$ is generated via gradient operators to serve as supervision for constructing a boundary-guided loss $\mathcal{L}_{\mathrm{edge}}$. A boundary attention map $B(x,y)$ is fused with the mask features through a Hadamard product operation. To further improve the model's understanding of true fan contours, a boundary structure alignment method was introduced based on support-query consistency matching. Given support features X_s and query features X_q, alignment is performed via Equation 4 as

$$\tilde{X}_q = \mathrm{SoftMax}\left(\frac{X_q W_q \cdot (X_s W_k)^T}{\sqrt{d_k}} \right) \cdot X_s W_v, \tag{4}$$

where W_q, W_k, and W_v are learnable parameters, and d_k is the scaling factor. This operation ensures localized consistency

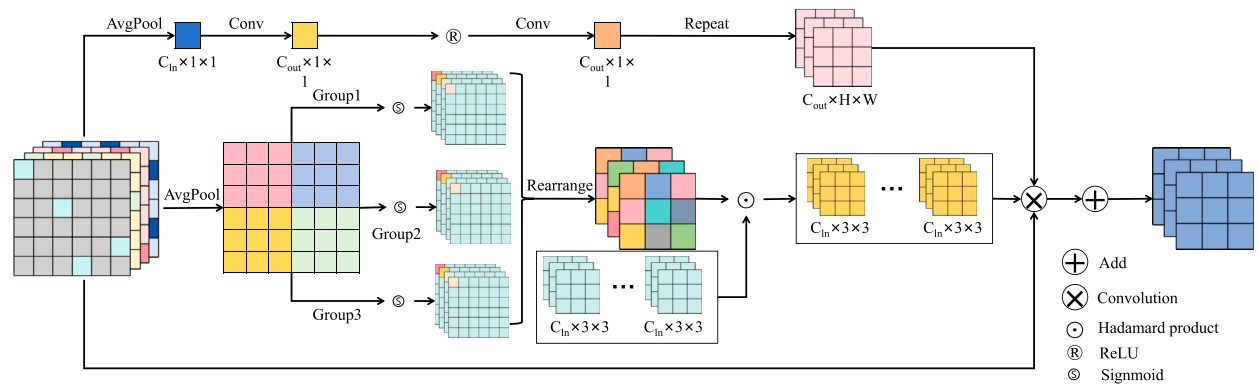

FIGURE 3
Illustration of the scale-adaptive optimization module. The module leverages grouped convolutions and channel-wise attention to dynamically adjust feature extraction at different scales. Adaptive weighting is achieved via a combination of average pooling, grouped convolution layers, and element-wise operations, allowing the model to enhance multi-scale perception for alluvial fan structures.

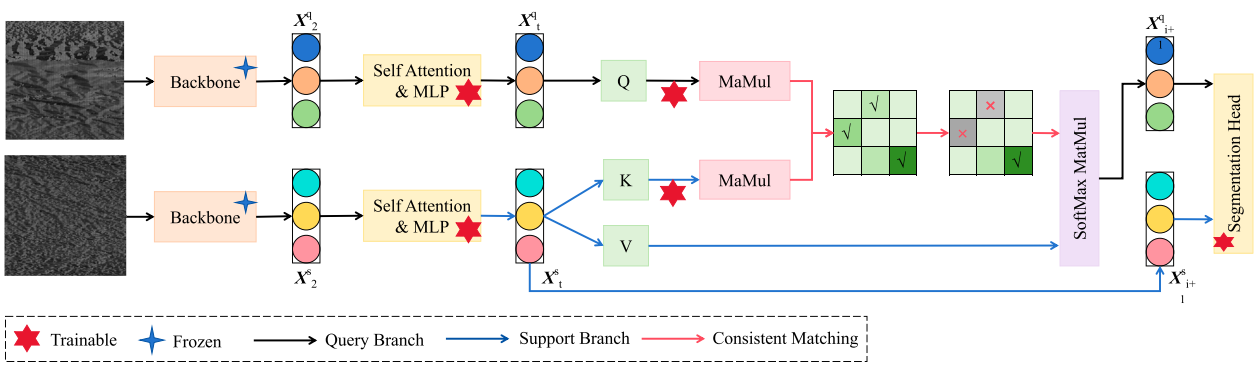

FIGURE 4
Illustration of the mask refinement and boundary-aware segmentation module. The architecture adopts a dual-branch structure, consisting of a trainable query branch and a frozen support branch. Through consistent matching between query and support features, fine-grained boundary representations are enhanced using self-attention and mask alignment mechanisms. The output is directed to a segmentation head for boundary-aware prediction.

mapping in feature space, improving boundary representation across varying scales and spatial positions. The loss function incorporates the original classification loss $\mathcal{L}_{\text{cls}}$, bounding box regression loss $\mathcal{L}_{\text{bbox}}$, and mask loss $\mathcal{L}_{\text{mask}}$, with the addition of the boundary-guided loss $\mathcal{L}_{\text{edge}}$. Equation 5 is defined as

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{cls}} + \lambda_2 \mathcal{L}_{\text{bbox}} + \lambda_3 \mathcal{L}_{\text{mask}} + \lambda_4 \mathcal{L}_{\text{edge}}, \tag{5}$$

where each weight λ_i is optimized on a validation set. This module substantially improves the model's ability to reconstruct fan boundaries while maintaining overall segmentation precision. Experimental results indicated significant improvements in mIoU and boundary accuracy across multiple test regions, especially in areas with abrupt slope transitions and overlapping fans. The module demonstrated enhanced stability, consistent boundary alignment, and reduced occurrences of mask drift and edge discontinuities, thereby improving both geomorphic segmentation accuracy and interpretability.

4 Experiment and results

4.1 Experimental settings

4.1.1 Experimental configuration

The proposed enhanced MASK R-CNN model was trained and evaluated for remote sensing-based identification of alluvial fans under a unified software and hardware environment. The experimental framework was implemented using TensorFlow 1.15 as the deep learning backend, with all neural network components and data processing pipelines developed in Python 3.6. Image preprocessing and augmentation were carried out using standard image libraries such as OpenCV and NumPy. Model training employed the Adam optimizer with an initial learning rate set to 0.0001. A step decay strategy was adopted, whereby the learning rate was halved every 10 epochs to improve training stability and convergence speed. The entire training process was executed over 100 epochs with a batch size of 4. The high-dimensional six-channel input was used to balance training efficiency and GPU

memory consumption. A composite weighted loss function was adopted, incorporating classification loss, bounding box regression loss, mask segmentation loss, and boundary structure loss. The respective weights were optimized using cross-validation. For dataset partitioning, the annotated alluvial fan samples were randomly divided into training, validation, and testing subsets with a ratio of 70%, 15%, and 15%, respectively. The training set was used to optimize network parameters, the validation set guided hyperparameter tuning and early stopping, while the independent test set was reserved for final performance evaluation. This partition ensured a balanced representation of different geomorphological conditions across subsets, thereby reducing overfitting and improving the reliability of performance assessment. All experiments were conducted on a high-performance server equipped with an NVIDIA Tesla V100 GPU (32 GB VRAM), an Intel Xeon Gold 6248 processor (2.50 GHz), and 256 GB of RAM. This configuration ensured sufficient training and inference efficiency for handling complex network structures and multi-scale input data.

4.1.2 Evaluation metrics

To comprehensively assess the performance of the proposed method in terms of both accuracy and boundary delineation, several quantitative metrics were employed. These included classification-based metrics such as Equations 6–9, as well as Equation 10 (mIoU) for evaluating segmentation consistency. In addition, visual analysis was conducted to qualitatively assess mask contour and boundary fitting performance. The metric definitions are given as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{8}$$

$$\text{F1} - \text{score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{9}$$

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i + FN_i}, \tag{10}$$

here, TP (true positive) denotes the number of correctly predicted pixels belonging to the fan region, TN (true negative) refers to correctly predicted non-fan pixels, FP (false positive) represents the number of pixels wrongly identified as fans, and FN (false negative) indicates missed fan pixels. mIoU, a widely used metric in semantic segmentation, computes the mean overlap between predicted and ground-truth areas across all classes. In addition to numerical metrics, qualitative comparisons were performed by overlaying predicted masks on ground-truth contours to visually assess the alignment of fan boundaries, particularly focusing on the model's performance in edge detail restoration.

4.1.3 Baseline comparisons

To verify the effectiveness of the proposed method, several mainstream approaches were selected as baselines, covering both traditional machine learning and modern deep learning architectures. These included the original three-channel MASK R-CNN (He et al., 2017), U-Net (Ronneberger et al., 2015), DeepLabv3+ (Peng et al., 2020), a random forest with spectral indices method (RF + Spectral Indices) (Boonprong et al., 2018), and the high-resolution network HRNet (Yu et al., 2021). Among them, the original MASK R-CNN served as the structural reference for evaluating the improvements introduced by topographic-spectral fusion and module enhancements. U-Net, known for its simplicity and strong edge preservation, is widely used in small-object segmentation in remote sensing. DeepLabv3+, utilizing atrous convolutions, enables multi-scale feature extraction suitable for large-scale geomorphological structures. The RF + Spectral Indices method, rooted in traditional remote sensing, relies on spectral and morphological descriptors with strong interpretability but limited generalization. HRNet, a recent state-of-the-art model, maintains high-resolution representations across multiple scales and excels in structural continuity and boundary detail preservation. These baselines represent diverse perspectives in current methodologies and provide a comprehensive benchmark for evaluating the performance gains achieved by the proposed framework.

4.2 Overall classification performance comparison of different models on alluvial fan recognition

This experiment was designed to evaluate the overall classification performance of the proposed six-channel enhanced MASK R-CNN, which integrates topographic and spectral information, for alluvial fan recognition by comparing it against several representative remote sensing models. To this end, five baseline models were selected, including a traditional machine learning approach (RF + spectral indices), classic semantic segmentation networks (U-Net, DeepLabv3+), a high-resolution representation network (HRNet), and the standard three-channel MASK R-CNN. A comprehensive performance assessment was conducted across four metrics: accuracy, precision, recall, and F1-score.

As shown in Table 2; Figures 5–7, the proposed method outperformed all comparison models across all metrics, particularly demonstrating superior capability in recall and F1-score. This confirms the effectiveness of the six-channel input and multi-module collaborative architecture in the task. From a theoretical standpoint, the RF + spectral indices approach relies on handcrafted features derived from spectral indices and lacks deep semantic modeling capacity, resulting in lower recall, especially in complex boundary regions of irregular fan shapes. U-Net and DeepLabv3+ adopt encoder-decoder architectures with reasonable feature reconstruction capabilities; however, their lack of topographic priors limits the ability to capture geometric structures of alluvial fans. HRNet maintains high-resolution representation through parallel multi-scale branches and exhibits certain advantages in continuous structure recognition. Nevertheless, it remains confined to two-dimensional spatial modeling and lacks the incorporation of three-dimensional features such as elevation and slope, which restricts its capability to model volumetric forms. The proposed method incorporates DEM-derived terrain bands at the input level and encodes six-channel data collaboratively via a ResNet backbone, while the scale-aware and boundary refinement modules

TABLE 2 Overall classification performance comparison of different models on alluvial fan recognition.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RF + Spectral Indices	78.6	74.3	69.5	71.8
U-Net	85.2	82.7	80.1	81.4
DeepLabv3+	86.9	84.5	82.0	83.2
HRNet	88.3	86.4	84.2	85.3
MASK R-CNN (3ch)	87.1	85.0	82.3	83.6
Proposed Method (6ch)	**91.7**	**89.8**	**88.5**	**89.1**

Bold values indicate the best performance.

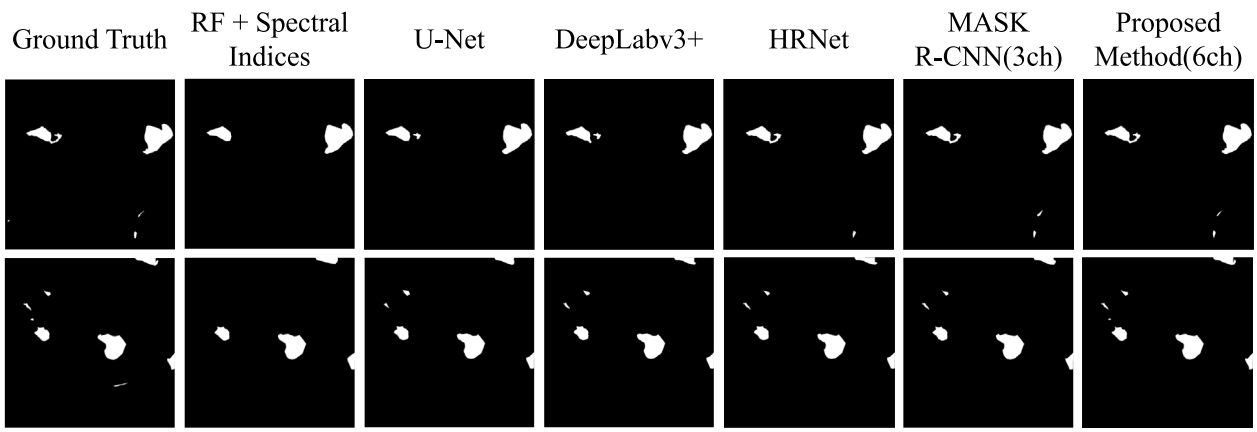

FIGURE 5
Visualization of segmentation on different methods.

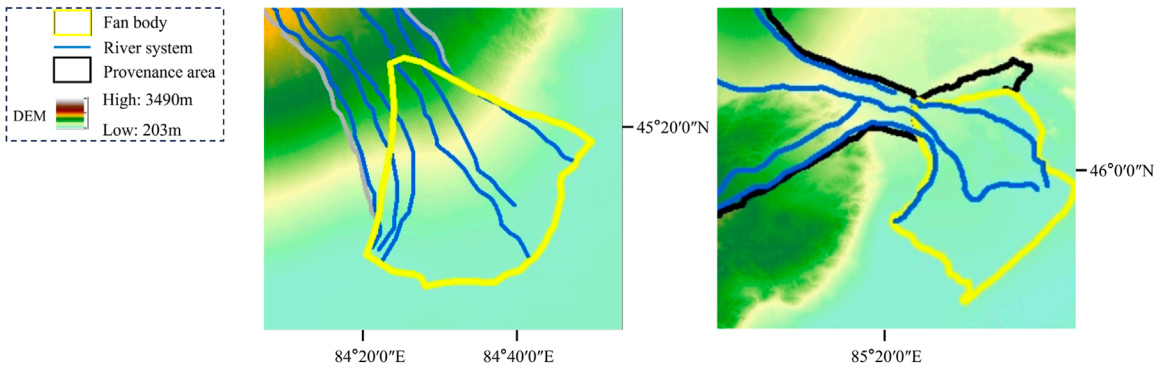

FIGURE 6
Overall boundary map of the Junggar Basin and the alluvial fan on the southern edge of the Qilian Mountains.

jointly enhance the network's ability to perceive and reconstruct multi-scale fan boundaries. Mathematically, the terrain-spectral joint input extends the feature space dimensionality, and the integration of attention mechanisms and mask optimization boosts high-frequency detail modeling, ultimately resulting in improved recognition accuracy and boundary recovery.

4.3 Ablation study on key modules of the proposed method

This experiment was conducted to systematically evaluate the contribution of three core modules in the proposed model—terrain-spectral fusion (TSF), scale-aware module (SAM), and mask

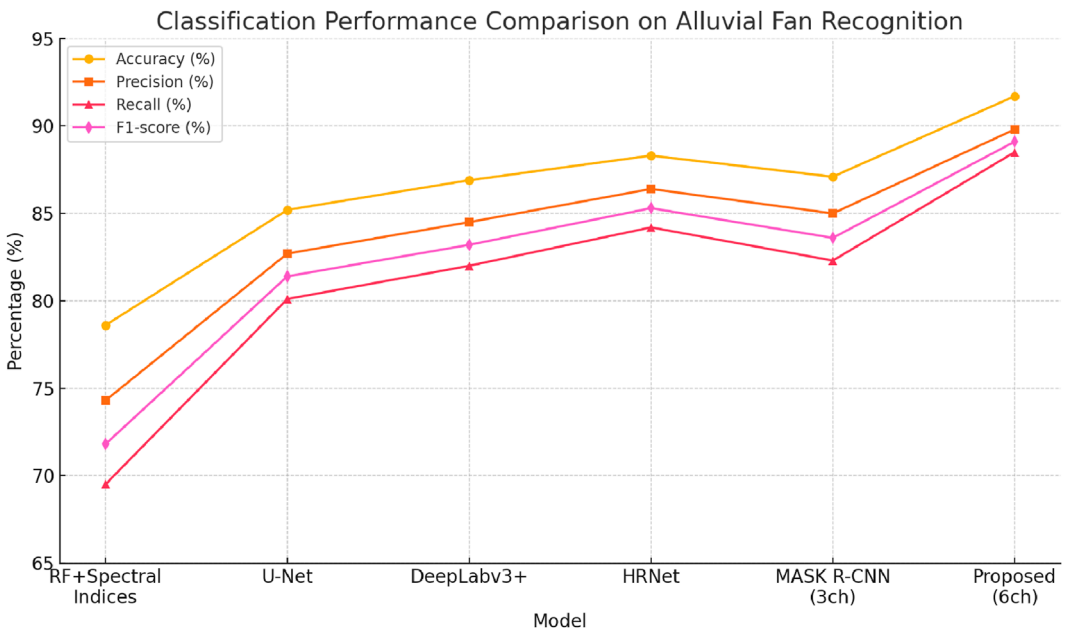

FIGURE 7
Overall classification performance comparison of different models on alluvial fan recognition.

boundary refinement (MBR)—through a stepwise ablation study. Using the three-channel MASK R-CNN as the baseline, each module was incrementally added, and the performance changes in terms of mIoU and boundary F1-score were recorded to determine the role of each component in enhancing geomorphological segmentation and boundary fitting.

As shown in Table 3; Figure 8, each module contributed significantly to performance improvement, with the full model achieving 81.5% mIoU and 80.4% boundary F1-score, representing increases of 5.7 and 6.6 percentage points over the baseline, respectively. The results highlight the advantage of the multi-module synergy. Theoretically, the TSF module enhances the model's ability to perceive geomorphic forms and spectral structures by introducing joint terrain-spectral feature encoding, effectively enriching the geometric and spectral priors in the feature space and improving semantic segmentation accuracy. The SAM module employs a local gradient-guided dynamic scale-aware mechanism to refine anchor distributions and feature weighting, allowing for adaptive modeling of complex fan structures across scales and improving robustness. The MBR module explicitly optimizes mask boundaries during decoding, using a support-query consistency alignment mechanism to enhance boundary recovery, which significantly boosts the boundary F1-score. From a mathematical modeling perspective, TSF extends the input tensor channel dimensionality, SAM introduces location-sensitive scale adjustment functions, and MBR applies an asymmetric attention mechanism to recalibrate edge representations in the mask. These innovations enable the model to better capture semantic structures and restore spatial details across different layers and spatial positions, thereby enhancing remote sensing recognition performance.

4.4 Performance consistency of the proposed method across fan size categories

This experiment was designed to validate the robustness and generalization ability of the proposed model across different scales of alluvial fans. The test dataset was categorized into three groups based on area: small (<0.5 km^2), medium (0.5–2 km^2), and large (>2 km^2). The model's performance was then evaluated using mIoU, boundary F1-score, and false negative rate to assess its stability under multi-scale and irregular geomorphic conditions.

As shown in Table 4; Figure 9, the proposed method maintained consistently high accuracy and boundary recovery across all categories, with an average mIoU of 81.5%, boundary F1-score of 80.4%, and a false negative rate of only 4.8%. Particularly strong performance was observed in medium and large fan regions, indicating the model's adaptability to complex fan structures. From the perspective of model architecture and mathematical characteristics, this consistency results from the synergy of the three core modules. The SAM module introduces a gradient-guided mechanism during multi-scale feature extraction, enabling the model to precisely capture primary radial and edge contours in large-scale fans, thereby enhancing mIoU. The MBR module strengthens boundary detail representation for small fans through boundary-aware optimization, mitigating issues of blurred boundaries and missed detections due to resolution limitations, thus improving boundary F1-score and reducing the false negative rate. The TSF module enriches the input feature representation by introducing spatial and spectral priors at the channel level, enabling differentiated modeling of fan morphology across scales from the encoding stage. Overall, the method exhibits strong scale invariance and structural consistency, making it

TABLE 3 Ablation study on key modules of the proposed method.

Model variant	TSF	SAM	MBR	mIoU (%)	Boundary F1 (%)
Baseline (3ch MASK R-CNN)	×	×	×	75.8	73.8
TSF	✓	×	×	78.3	74.9
TSF + SAM	✓	✓	×	80.1	77.0
TSF + MBR	✓	×	✓	79.5	78.6
SAM + MBR	×	✓	✓	77.2	76.9
Full Model	✓	✓	✓	**81.5**	**80.4**

Bold values indicate the best performance.

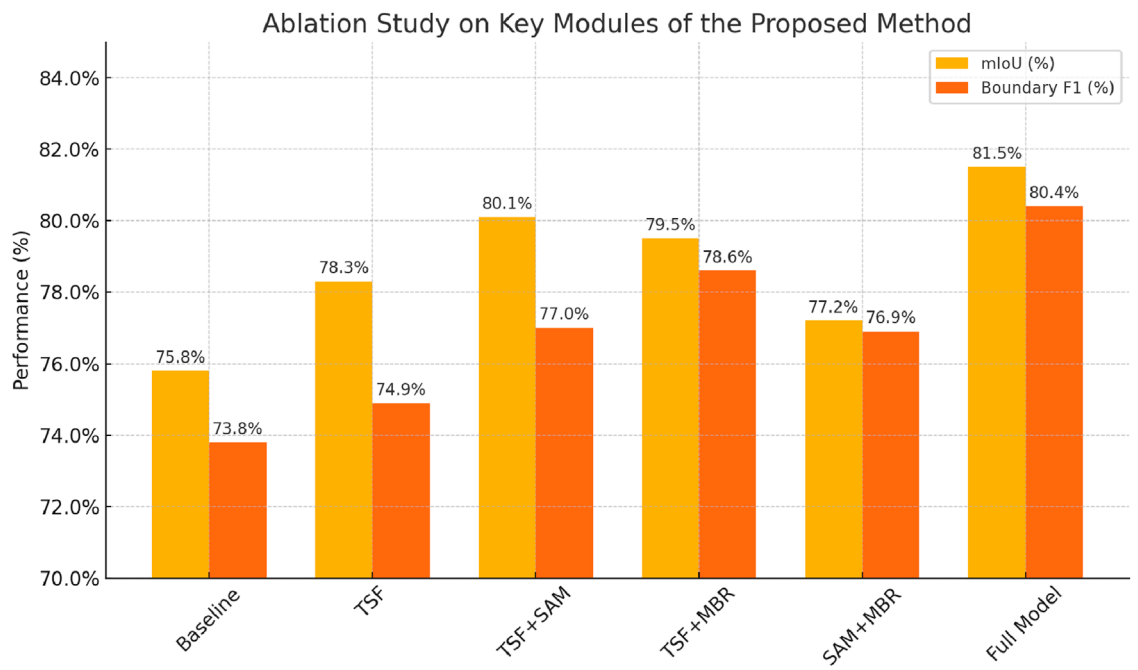

FIGURE 8
It illustrates the impact of different module combinations on mIoU and boundary F1 performance.

well-suited for automated recognition tasks involving diverse alluvial fan types.

5 Discussion

5.1 Practical value and applicability

The proposed remote sensing recognition method for alluvial fans, which integrates terrain-spectral features with multi-module optimization mechanisms, demonstrates substantial practical application value, particularly in typical geomorphological environments such as arid, semi-arid regions, and mountainous forelands. In real-world applications, alluvial fan areas are frequently associated with sudden natural hazards including flash floods, debris flows, and soil erosion. Traditional remote sensing methods relying on manual interpretation and rule-based extraction often struggle with the complex boundary morphologies and the coexistence of multi-scale geomorphic units, leading to high false detection rates and discontinuous delineation. Recent deep learning approaches such as U-Net, DeepLabv3+, HRNet, and the standard three-channel Mask R-CNN have improved recognition efficiency and automation; however, they still exhibit limitations in geomorphological tasks. For instance, U-Net and DeepLabv3+ rely primarily on spectral information and lack topographic priors, making them prone to blurred boundaries and misclassification in areas with complex terrain. HRNet maintains high-resolution features but remains constrained to two-dimensional modeling, failing to capture volumetric structures. The original three-channel Mask R-CNN offers stronger boundary localization but cannot

TABLE 4 Performance consistency of the proposed method across fan size categories.

Fan size category	mIoU (%)	Boundary F1 (%)	False negative rate (%)
Small Fans (<0.5 km^2)	76.9	79.8	6.1
Medium Fans (0.5–2 km^2)	82.1	81.3	4.3
Large Fans (>2 km^2)	83.2	80.1	3.9
Average	**81.5**	**80.4**	**4.8**

Bold values indicate the best performance.

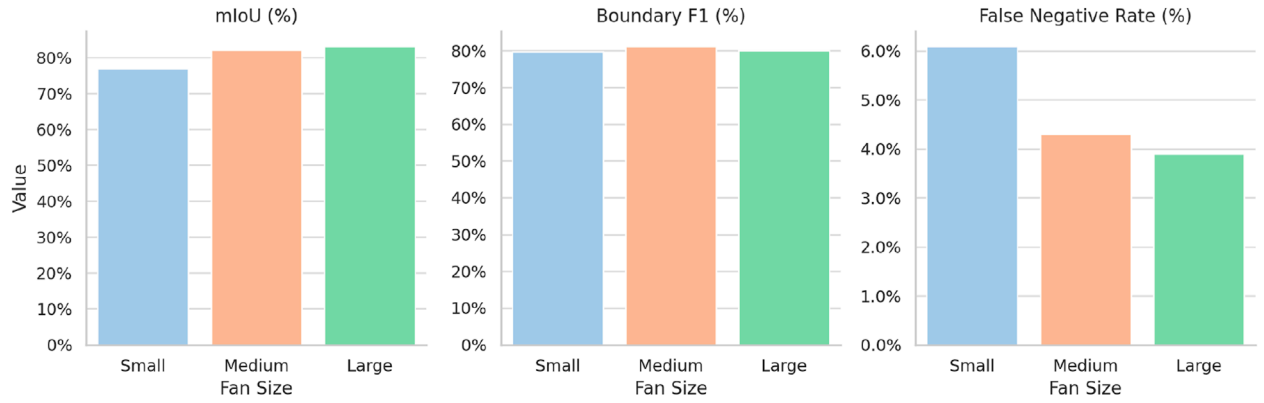

FIGURE 9
This figure presents the performance of the proposed method across different fan size categories (small, medium, large) in terms of mIoU, Boundary F1, and False Negative Rate.

effectively adapt to multi-scale fan morphologies due to the absence of terrain-spectral integration and boundary refinement. The presented model addresses these challenges by incorporating DEM-derived variables such as elevation, slope, and aspect, and introducing multi-scale attention modulation and boundary-guided refinement mechanisms, thereby enabling precise morphological modeling and detailed boundary reconstruction of alluvial fans. For example, in regions such as Xinjiang, Qinghai, and Gansu, where terrain undulations are pronounced and alluvial fans are extensively distributed across piedmont basins and river outlets, accurate recognition of fan morphology forms a critical foundation for flood risk assessment and territorial spatial planning. When processing large-scale high-resolution remote sensing imagery, the proposed method balances the representation of global fan distributions and the structural details of local lobes, providing technical support for the construction and dynamic updating of regional fan databases. Furthermore, in fields such as water resource assessment, agricultural irrigation planning, and ecological barrier construction, accurate delineation of alluvial fan boundaries and areal extents can assist in the demarcation of irrigation districts, identification of groundwater recharge zones, and the layout of desertification control projects. Therefore, compared with prior deep learning approaches, the improved Mask R-CNN model proposed in this study not only introduces methodological innovations from a theoretical perspective but also demonstrates stronger applicability and scalability in real-world geomorphological recognition tasks, particularly suitable for automated geoinformation extraction in environmentally fragile regions.

5.2 Limitation and future work

Despite the superior performance of the proposed multi-module recognition framework that integrates terrain-spectral features, particularly in terms of accuracy and boundary resolution, several limitations remain. First, the model still suffers from omission errors when detecting small-scale alluvial fans, especially in cases where fan boundaries are blurred or exhibit strong visual similarity to adjacent fans, suggesting that spatial feature modeling requires further refinement. Second, the generalization capability of the framework across diverse geomorphic regions has yet to be fully validated. In monsoon or humid environments, for example, where alluvial–colluvial composite fans are common, the method may face challenges due to limited spectral contrast and weak topographic variation. Third, the framework is highly dependent on high-quality DEM and high-resolution remote sensing imagery. In regions with limited multisource data availability, or where imagery is affected by occlusions and shadows, performance degradation is likely. Future research will therefore focus on improving detection performance for small fans and boundary-ambiguous areas, potentially through

the integration of higher-dimensional auxiliary data such as SAR radar or geological mapping units. Such enhancements would enable a more comprehensive perception of morphological details and genesis-related context. In addition, advancing cross-regional transferability and few-shot adaptation capabilities will be critical to strengthen robustness and generalizability in diverse environments. Ultimately, the aim is to build a continuously updateable, multi-geomorphology-adaptive automated recognition system for alluvial fans, providing sustained support for regional disaster assessment, resource management, and ecological development.

6 Conclusion

Alluvial fans, as typical geomorphic units in arid and semi-arid regions, play a pivotal role in disaster risk assessment, water resource management, and land-use planning. However, their complex spatial morphology and pronounced scale variability present significant challenges to traditional remote sensing techniques, often resulting in limited accuracy and imprecise boundary delineation. To overcome these limitations, this study proposes an enhanced Mask R-CNN framework that incorporates terrain-spectral features. The end-to-end architecture integrates a topographic-spectral fusion (TSF) input module, a scale-adaptive optimization module, and a mask-boundary refinement (MBR) module, collectively aimed at achieving high-precision recognition and fine-grained boundary characterization of alluvial fans.In comparative experiments against several mainstream methods, the proposed model achieved an accuracy of 91.7%, a precision of 89.8%, and a recall of 88.5%, with an F1-score reaching 89.1%, significantly outperforming conventional approaches such as Mask R-CNN, U-Net, and DeepLabv3+. Segmentation performance further demonstrated superiority, with a mean intersection over union (mIoU) of 81.5% and a boundary F1-score of 80.4%, highlighting strong capabilities in spatial structure modeling and edge delineation. Ablation studies confirmed the significant contributions of each core module, particularly the TSF and MBR components, in enhancing boundary detail representation. Additionally, scale-consistency analysis validated the model's robustness and stability across various fan size categories, especially in reducing omission rates for small-scale fans.Beyond its technical contributions, the proposed method offers foundational value for sedimentological research and facies analysis. By accurately capturing the morphological complexity and depositional patterns of alluvial fans, it facilitates improved understanding of sedimentary processes, facies associations, and alluvial architecture. Furthermore, the ability to extract detailed geomorphic and structural features supports paleogeographic reconstruction efforts—such as paleo-topography modeling, provenance analysis, and paleoclimate inference—by providing reliable spatial constraints and high-resolution terrain proxies. Overall, this research presents a generalized and transferable solution for the automated recognition of complex geomorphic systems, with broad applicability in intelligent remote sensing interpretation and terrain analysis in arid environments, while offering critical analog references for basin evolution studies and sedimentary system modeling.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

HZ: Software, Conceptualization, Writing – original draft, Methodology. SL: Project administration, Conceptualization, Writing – original draft, Funding acquisition. CZ: Writing – original draft, Conceptualization, Software, Methodology. ZM: Data curation, Writing – original draft, Visualization, Resources.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by National Natural Science Foundation of China grant number 61202479.

Conflict of interest

Authors HZ and ZM were employed by PetroChina. Author CZ was employed by Power China Urban Planning and Design Institute Co., Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Babič, M., Petrovič, D., Sodnik, J., Soldo, B., Komac, M., Chernieva, O., et al. (2021). Modeling and classification of alluvial fans with dems and machine learning methods: a case study of slovenian torrential fans. *Remote Sens.* 13, 1711. doi:10.3390/rs13091711

Boonprong, S., Cao, C., Chen, W., and Bao, S. (2018). Random forest variable importance spectral indices scheme for burnt forest recovery monitoring—multilevel rf-vimp. *Remote Sens.* 10, 807. doi:10.3390/rs10060807

Gao, B.-C. (1996). Ndwi—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. doi:10.1016/s0034-4257(96)00067-3

Ghahraman, K. (2024). Comprehensive study of alluvial fans: geomorphology, hazard assessment, and anthropogenic interactions in arid and semi-arid environments of Iran.

Ghahraman, K., and Nagy, B. (2023). Flood risk on arid alluvial fans: a case study in the joghatay mountains, northeast Iran. *J. Mt. Sci.* 20, 1183–1200. doi:10.1007/s11629-022-7635-8

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn", in *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

Hou, T., and Li, J. (2024). Application of mask r-cnn for building detection in uav remote sensing images. *Heliyon* 10, e38141. doi:10.1016/j.heliyon.2024.e38141

Ismael, R. Q., and Sadeq, H. A. (2025). Sequential hybrid integration of u-net and fully convolutional networks with mask r-cnn for enhanced building boundary segmentation from satellite imagery. *Zanco J. Pure Appl. Sci.* 37, 157–171. doi:10.21271/zjpas.37.3.13

Jiang, Y., Si, C., and Yang, L. (2024). "Improvement strategies for mask r-cnn in satellite image analysis", in *2024 3rd international conference on electronics and information technology (EIT)* (IEEE), 739–744.

Li, S., Xiong, L., Tang, G., and Strobl, J. (2020). Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology* 354, 107045. doi:10.1016/j.geomorph.2020.107045

Li, J., Cai, Y., Li, Q., Kou, M., and Zhang, T. (2024). A review of remote sensing image segmentation by deep learning methods. *Int. J. Digital Earth* 17, 2328827. doi:10.1080/17538947.2024.2328827

Lin, X., Wa, S., Zhang, Y., and Ma, Q. (2022). A dilated segmentation network with the morphological correction method in farming area image series. *Remote Sens.* 14, 1771. doi:10.3390/rs14081771

Lv, J., Shen, Q., Lv, M., Li, Y., Shi, L., and Zhang, P. (2023). Deep learning-based semantic segmentation of remote sensing images: a review. *Front. Ecol. Evol.* 11, 1201125. doi:10.3389/fevo.2023.1201125

Lyu, P., He, L., He, Z., Liu, Y., Deng, H., Qu, R., et al. (2021). Research on remote sensing prospecting technology based on multi-source data fusion in deep-cutting areas. *Ore Geol. Rev.* 138, 104359. doi:10.1016/j.oregeorev.2021.104359

Mei, S., Lian, J., Wang, X., Su, Y., Ma, M., and Chau, L.-P. (2024). A comprehensive study on the robustness of deep learning-based image classification and object detection in remote sensing: surveying and benchmarking. *J. Remote Sens.* 4, 0219. doi:10.34133/remotesensing.0219

Miliaresis, G. C., and Argialas, D. (2000). Extraction and delineation of alluvial fans from digital elevation models and landsat thematic mapper images. *Photogrammetric Eng. Remote Sens.* 66, 1093–1101.

Odunuga, S., and Raji, S. (2018). Geomorphological mapping of part of the niger delta, Nigeria using dem and multispectral imagery. *J. Nat. Sci. Eng. Technol.* 17, 121–146. doi:10.51406/jnset.v17i1.1904

Peng, H., Xue, C., Shao, Y., Chen, K., Xiong, J., Xie, Z., et al. (2020). Semantic segmentation of litchi branches using deeplabv3+ model. *Ieee Access* 8, 164546–164555. doi:10.1109/access.2020.3021739

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer), 234–241.

Shoshta, A., and Marh, B. S. (2023). Alluvial fans of trans-himalayan cold desert (pin valley, India): quantitative morphology and controlling factors. *Phys. Geogr.* 44, 136–161. doi:10.1080/02723646.2021.1907883

Thannoun, R. G., Beti, A. K., and Al-Sa'igh, L. K. (2016). Identifying alluvial fans features using multispectral image processing techniques in selected area, northern Iraq. *Sulaimani J. Pure Appl. Sci.* 18, 133–146.

Udin, W., Norazami, N., Sulaiman, N., Zaudin, N. C., Ma'ail, S., and Nor, A. M. (2019). "Uav based multi-spectral imaging system for mapping landslide risk area along jeli-gerik highway, jeli, kelantan," in *2019 IEEE 15th international colloquium on signal processing & its applications (CSPA)* (IEEE), 162–167.

Wan, C., Gan, J., Chen, A., Acharya, P., Li, F., Yu, W., et al. (2024). A novel method for identifying landslide surface deformation *via* the integrated yolox and mask r-cnn model. *Int. J. Comput. Intell. Syst.* 17, 255. doi:10.1007/s44196-024-00655-w

Wang, H., and Li, X. (2024). Expanding horizons: U-net enhancements for semantic segmentation, forecasting, and super-resolution in ocean remote sensing. *J. Remote Sens.* 4, 0196. doi:10.34133/remotesensing.0196

Wang, Y., Yang, L., Liu, X., and Yan, P. (2024). An improved semantic segmentation algorithm for high-resolution remote sensing images based on deeplabv3+. *Sci. Rep.* 14, 9716. doi:10.1038/s41598-024-60375-1

Yu, C., Xiao, B., Gao, C., Yuan, L., Zhang, L., Sang, N., et al. (2021). "Lite-hrnet: a lightweight high-resolution network," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10440–10450.

Zhang, L., Zhang, Y., and Ma, X. (2021). "A new strategy for tuning relus: Self-adaptive linear units (salus)," in *ICMLCA 2021; 2nd international conference on machine learning and computer application* (Shenyang, China: VDE), 1–8.

Zhang, Y., Wang, H., Xu, R., Yang, X., Wang, Y., and Liu, Y. (2022). High-precision seedling detection model based on multi-activation layer and depth-separable convolution using images acquired by drones. *Drones* 6, 152. doi:10.3390/drones6060152

Zhou, X., Chen, S., Ren, Y., Zhang, Y., Fu, J., Fan, D., et al. (2022). Atrous pyramid gan segmentation network for fish images with high performance. *Electronics* 11, 911. doi:10.3390/electronics11060911