remote sensing

MDPI

Article

GCF-Net: A Geometric Context and Frequency Domain Fusion Network for Landslide Segmentation in Remote Sensing Imagery

Chunlong Du [1,2], Shaoqun Qi [1,2,*], Luhe Wan [1,2], Yin Chen [1,2], Zhiwei Lin [1,2], Ling Zhu [1,2] and Xiaona Yu [1,2]

[1] College of Geographical Sciences, Harbin Normal University, Harbin 150025, China; chunlongdu@stu.hrbnu.edu.cn (C.D.); wanluhe@163.com (L.W.); chenyin@stu.hrbnu.edu.cn (Y.C.); linzhiwei@stu.hrbnu.edu.cn (Z.L.); zling@stu.hrbnu.edu.cn (L.Z.); yuxiaon@stu.hrbnu.edu.cn (X.Y.)
[2] Heilongjiang Wuyiling Wetland Ecosystem National Observation and Research Station, Yichun 153000, China
* Correspondence: qishaoqun@hrbnu.edu.cn

Highlights

What are the main findings?

- GCF-Net is proposed, integrating frequency-domain enhancement with multi-scale spatial context for landslide boundary segmentation, achieving OA of 96.42% on the Hybrid dataset and 96.71% on the LMHLD dataset.
- Experimental results demonstrate that the three-stage framework progressively refines features, significantly improving edge precision and maintaining global semantic integrity.

What are the implications of the main findings?

- Provides a high-precision framework for remote sensing-based landslide detection, supporting disaster assessment and post-event rescue in complex terrains.
- Validates the effectiveness of frequency–spatial fusion strategies in preserving boundary details while suppressing noise, offering methodological reference for future edge-aware segmentation research.

Abstract

Remote sensing-based landslide segmentation is of great significance for geological hazard assessment and post-disaster rescue. Existing convolutional neural network methods, constrained by the inherent limitations of spatial convolution, tend to lose high-frequency edge details during deep semantic extraction, while frequency-domain analysis, although capable of globally preserving high-frequency components, struggles to perceive local multi-scale features. The lack of an effective synergistic mechanism between them makes it difficult for networks to balance regional integrity and boundary precision. To address these issues, this paper proposes the Geometric Context and Frequency Domain Fusion Network (GCF-Net), which achieves explicit edge enhancement through a three-stage progressive framework. First, the Pyramid Lightweight Fusion (PGF) block is proposed to aggregate multi-scale context and provide rich hierarchical features for subsequent stages. Second, the Geometric Context and Frequency Domain Fusion (GCF) module is designed, where the frequency-domain branch generates dynamic high-frequency masks via the Fourier transform to locate boundary positions, while the spatial branch models foreground–background relationships to understand boundary semantics, with both branches fused through an adaptive gating mechanism. Finally, Edge-aware Detail Consistency Improvement (EDCI) module is designed to balance boundary preservation and noise suppression based on edge confidence, achieving adaptive output refinement. Under the joint supervision of

Check for updates

Academic Editor: Antonio Miguel Ruiz Armenteros

Received: 31 December 2025
Revised: 4 February 2026
Accepted: 16 February 2026
Published: 18 February 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.

Focal loss, Dice loss, and Edge loss, experiments on the mixed dataset and LMHLD dataset demonstrate that GCF-Net achieves OAs of 96.42% and 96.71%, respectively. Ablation experiments and visualization results further validate the effectiveness of each module and the significant improvement in boundary segmentation.

Keywords: landslide segmentation; remote sensing imagery; frequency-domain fusion; edge-aware segmentation

1. Introduction

Landslides are among the most destructive geological hazards worldwide, causing substantial casualties and economic losses annually due to their sudden onset and widespread distribution [1–3]. Accurate landslide monitoring is therefore critical for disaster assessment, emergency response, and early warning, particularly in complex terrain. With the rapid advancement of Earth observation technology, high-resolution remote sensing imagery has become the primary approach for landslide investigation, offering distinct advantages in terms of wide-area coverage, real-time acquisition, and non-contact detection [4,5]. However, achieving automated and high-precision landslide segmentation from large-scale high-resolution imagery remains a significant challenge. The advent of deep learning has revolutionized automatic landslide segmentation. U-Net [6] achieved multi-scale feature fusion through its encoder-decoder structure and skip connections, while DeepLabV3+ [7] enhanced context aggregation using the Atrous Spatial Pyramid Pooling (ASPP) module; both have established themselves as foundational architectures for remote sensing semantic segmentation. However, the inherent local connectivity of convolution operations limits the effective receptive field. This constraint makes it difficult for networks to capture the holistic structure of large-scale landslides, often leading to segmentation results characterized by internal fragmentation or regional discontinuity [8,9].

To circumvent the limitations of local receptive fields, researchers have explored architectures with explicit global context modeling. SegFormer [10] and SwinUnet [11] leverage self-attention mechanisms to capture long-range dependencies. Concurrently, recent approaches based on State Space Models (SSMs), notably RS3Mamba [12] and SegMAN [13], have integrated SSM blocks to achieve efficient long-range modeling with linear complexity. Despite significantly improving global semantic consistency, these methods face a common bottleneck in fine-grained boundary segmentation. A critical limitation stems from the spatial structure simplification inherent in their global mechanisms, such as patch partitioning in Transformers or sequence serialization in RS3Mamba [12] and SegMAN [13]. Such operations inevitably compromise effective resolution or disrupt precise local continuity. Consequently, feature representations become overly smoothed, attenuating the high-frequency details essential for defining landslide boundaries, particularly in the complex transition zones between landslides and their surrounding backgrounds.

The attenuation of high-frequency information is particularly critical in the context of landslide remote sensing imagery [14]. Landslide boundaries are typically characterized by gradual transitions into backgrounds such as vegetation and bare land, making the preservation of high-frequency details indispensable for accurate delineation. Once these components are suppressed during feature extraction, boundary regions become prone to confusion with the background, leading to over-smoothed segmentation or boundary displacement [15,16]. To address this, existing research has primarily sought solutions within the spatial domain, yet with limited success. G-CASCADE [17] explicitly targets boundary refinement through cascading aggregation modules; however, its effectiveness relies

heavily on the spatial quality of intermediate features, which struggles to recover precise geometric contours when input representations are already smoothed by deep encoders. Similarly, DAE-Former [18] utilizes diagonal attention mechanisms to enhance global context modeling, but this essentially functions as a similarity-based feature smoothing process, often causing weak-textured landslide edges to be overwhelmed by the dominant background. HRNet [19] maintains high-resolution representations through parallel branches, but by relying solely on multi-scale feature stacking in the spatial dimension, it remains ineffective at distinguishing highly similar textures in complex transition zones. Fundamentally, these methods rely on the reorganization and reweighting of spatial pixels. When original high-frequency details are lost in deep networks, purely spatial-domain operations are inherently limited in recovering geometric details, thereby constraining further improvements in segmentation accuracy [20,21].

As spatial-domain representations approach their performance bottlenecks, frequency-domain analysis offers a novel perspective for mitigating boundary blurring. Based on the Fourier transform, images can be decomposed into low-frequency components representing global structure and high-frequency components capturing geometric details, such as edges and textures. This spectral property inherently aligns with the dual objectives of landslide segmentation: preserving semantic integrity while recovering precise boundary details. Consequently, various frequency-domain methods have been introduced into computer vision tasks. However, existing approaches exhibit distinct limitations when applied to landslide segmentation. For instance, FcaNet [22] extends channel attention to the frequency domain via the discrete cosine transform; while this enhances frequency perception, it compresses spatial information into channel descriptors, effectively neglecting spatial high-frequency distributions and failing to recover pixel-level boundaries. Similarly, Fast Fourier Convolution (FFC) [23] utilizes global frequency-domain convolution to expand the receptive field, but it applies indiscriminate processing to both high and low frequencies, lacking the targeted enhancement required for boundary-critical details. Furthermore, Octave Convolution [24] partitions features into frequency groups using fixed ratios, but this rigid decomposition fails to adapt to the heterogeneity of landslide boundaries, which range from distinct rock cross-sections to blurred vegetation interfaces. Finally, FreqFusion [25] introduces frequency alignment to alleviate multi-scale mismatches, but it prioritizes the smooth transition of low-frequency structures rather than the explicit enhancement of high-frequency edges.

To surmount the limitations of single-domain approaches, recent studies have sought to construct hybrid architectures that synergize spatial and frequency-domain representations. Notably, SFFNet [26] proposes a wavelet-based fusion network tailored for remote sensing, utilizing the Haar wavelet transform to decompose features and enhance detail recovery via multi-domain integration. Similarly, WaveViT [27] incorporates wavelet transforms into Vision Transformers to enable invertible down-sampling, preserving structural and textural details typically compromised by standard attention-based frameworks. From a Fourier perspective, GFNet [28] employs a global filter network to learn long-range dependencies directly in the frequency domain. However, these methods typically apply indiscriminate enhancement to high-frequency components, failing to account for the inherent complexity of landslide scenes. In landslide imagery, the high-frequency spectrum contains both salient boundary cues and irrelevant background textures. Absent a mechanism to differentiate between these elements, the uniform amplification of high-frequency signals inevitably boosts noise alongside target edges, thereby obscuring the distinction between landslide features and the background in complex transition zones.

Beyond the challenges associated with fusion strategies, current frequency-based and hybrid approaches encounter two additional architectural limitations. First, multi-scale

feature integration is often inadequate. In landslide detection, small-scale landslides are primarily characterized by high-frequency components, whereas large-scale landslides rely on low-frequency structural information [29]. Enhancement performed solely on a single scale fails to simultaneously accommodate the boundary features of targets varying in size, underscoring the necessity of robust multi-scale representations as a foundation for frequency analysis [30,31]. Second, post-enhancement refinement is frequently neglected. While high-frequency injection improves details, it inevitably introduces residual noise. However, existing architectures typically lack a dedicated stage to suppress these potential disturbances, which can compromise the precision of the optimized boundaries.

To address the aforementioned limitations, this paper proposes the Geometric Context and Frequency Domain Fusion Network (GCF-Net), designed to overcome the accuracy bottlenecks of landslide boundary segmentation in complex terrains by deeply synergizing frequency-domain analysis with multi-scale spatial semantics. Unlike existing methods that rely on fixed frequency processing or single-dimensional enhancement, GCF-Net adaptively integrates frequency-domain geometric features with spatial contextual information, thereby achieving precise boundary localization while preserving global semantic integrity. Specifically, the architecture comprises three core modules. First, in the feature aggregation stage, the Pyramid Lightweight Fusion (PGF) block combines lightweight convolution with pyramid pooling to generate compact multi-scale representations, providing a hierarchical semantic foundation for detecting landslides of varying scales. Second, the Geometric Context and Frequency Domain Fusion (GCF) module serves as the core engine in the edge perception stage. This module employs a frequency branch to generate dynamic high-frequency masks for adaptive boundary separation, and a spatial branch to model foreground–background semantic relationships. The two branches are then fused via a gating mechanism to output edge-enhanced features. Third, for the detail refinement stage, the Edge-aware Detail Consistency Improvement (EDCI) module is utilized. Guided by the edge-enhanced features from the GCF, this module uses edge confidence maps to preserve fine details in high-confidence boundary regions while suppressing noise in low-confidence background areas, ensuring adaptive optimization of the frequency-enhanced results. Collectively, these stages establish a complementary and progressive framework where feature representations are iteratively refined to achieve accurate boundary segmentation alongside global semantic coherence. Furthermore, we introduce a joint supervision strategy employing Focal Loss, Dice Loss, and Edge Loss to constrain the network across three dimensions: pixel classification, regional overlap, and boundary alignment.

1. To effectively balance regional integrity and boundary precision in segmentation tasks, the Geometric Context and Frequency Domain Fusion Network (GCF-Net) is constructed. It explicitly enhances edge information through collaborative modeling of frequency and spatial domains, capturing high-frequency boundary details while maintaining a low-complexity architecture.
2. To overcome feature redundancy and achieve effective architectural optimization, the Pyramid Lightweight Fusion (PGF) block is constructed. It significantly reduces computational complexity and parameter overhead by combining lightweight convolution with pyramid pooling, thereby providing a highly efficient feature foundation for subsequent frequency-spatial modeling.
3. To mitigate blurred boundaries and weakened textures in landslide imagery, the Geometric Context and Frequency Domain Fusion (GCF) module is constructed, which integrates dynamic high-frequency masks from the frequency branch with foreground–background semantics from the spatial branch via an adaptive gating mechanism, thereby enhancing feature discriminability in boundary regions.

4. To mitigate small target omission and edge discontinuity, the Edge-aware Detail Consistency Improvement (EDCI) module is constructed, which adaptively balances boundary preservation and noise suppression through boundary confidence guidance, thereby enhancing the geometric precision and semantic coherence of the segmentation results.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed network architecture. Section 4 describes the experiments and provides data analysis. Section 5 offers discussion, and Section 6 concludes the paper.

2. Related Works

2.1. Semantic Segmentation Networks for Landslide Detection

Deep learning-driven semantic segmentation is the dominant paradigm for landslide detection in remote sensing [32,33]. Early research predominantly leveraged CNN-based encoder-decoder architectures. For instance, PSPNet [34] uses pyramid pooling to capture multi-scale global features, while BiSeNet [35,36] balances semantics and details via dual-path structures for real-time tasks. Similarly, EfficientNet optimizes the accuracy-efficiency trade-off through compound scaling strategies. To further enhance representation, architectures like CBAM [37] and SKNet [38] integrate attention mechanisms and selective kernels, respectively, enabling networks to adaptively focus on discriminative regions and adjust receptive fields for multi-scale targets.

To address the inherent locality constraints of convolution operations, Vision Transformers have been progressively incorporated into the segmentation domain. PVTv2 [39] mitigates the computational overhead of self-attention through pyramid-style multi-scale feature extraction, enabling the processing of high-resolution remote sensing imagery. PoolFormer [40] replaces complex self-attention computations with simple pooling operations, validating the efficacy of meta-architecture designs. SegNeXt [41] revisits the potential of convolutional attention, achieving performance comparable to Transformers via multi-scale convolutional aggregation. Furthermore, TopFormer [42] explores hybrid architectures combining CNNs and Transformers to balance local detail capture with global dependency modeling. Recently, state space models have garnered significant attention, with VMamba [43] introducing the Mamba architecture to visual tasks, achieving long-range sequence modeling with linear complexity.

It is evident that existing state-of-the-art architectures primarily focus on enhancing global semantic coherence through complex spatial interaction mechanisms, yet often neglect the independent representation and explicit preservation of high-frequency geometric details. In contrast, our proposed method pioneers a synergistic approach combining frequency and spatial domains. By leveraging Fourier analysis to explicitly localize and recover the spectral responses of landslide boundaries, our method precisely reconstructs geometric contours within deep semantic features.

2.2. Frequency Domain Learning in Visual Recognition

Frequency-domain analysis decomposes images into low-frequency structures and high-frequency details via the Fourier transform, offering representations complementary to the spatial domain. This perspective is particularly valuable for boundary-sensitive segmentation tasks. In the realm of global feature modeling, GFNet [28] employs learnable frequency-domain filters to replace self-attention mechanisms, thereby maintaining global modeling capabilities while significantly reducing computational complexity. AFNO [44] advances this by developing adaptive Fourier neural operators, demonstrating an excellent balance between efficiency and accuracy in large-scale visual tasks. Similarly, Flatten

Transformer [45] offers a frequency-domain interpretation of linear attention mechanisms, unveiling the intrinsic links between attention computation and spectral operations.

In the context of multi-resolution representation and image restoration, WaveMLP [46] incorporates wavelet transforms into MLP architectures, enhancing hierarchical representation through multi-resolution decomposition. SFNet [47] introduces semantic flow alignment, leveraging frequency-domain properties to guide efficient multi-scale fusion. Furthermore, Restormer [48] integrates Transformers with frequency operations for efficient global modeling, while SwinIR [49] exploits the window attention of Swin Transformer to improve detail recovery. More recently, FreeU [50] has shown effective improvement in generative model output by suppressing high-frequency noise in decoders through frequency-domain reweighting.

While effective, these methods are predominantly tailored for classification and restoration, failing to address the stringent boundary precision required for landslide segmentation. Furthermore, their reliance on static filtering lacks the flexibility to adapt to scene content, making it difficult to distinguish true boundary signals from background noise. To bridge this gap, we propose a deep synergistic strategy integrating frequency and spatial semantics, enabling the adaptive extraction of boundary-specific spectral responses.

2.3. Boundary-Aware Segmentation Approaches

Boundary precision is a pivotal metric for assessing semantic segmentation quality. Consequently, researchers have extensively explored boundary-aware strategies, primarily through two distinct avenues: dual-stream architectures and multi-scale fusion. Regarding dual-stream architectures, PointRend [51] conceptualizes segmentation as a rendering problem, enhancing contour precision via adaptive point sampling and iterative refinement in uncertain regions. SegFix [52] addresses boundary misalignment by performing offset correction during post-processing. InverseForm [53] introduces a framework for boundary inversion learning to explicitly model boundary transformations. Furthermore, BoundaryFormer [54] formulates boundary detection as a sequence prediction task, leveraging Transformers for end-to-end optimization. PIDNet [55] employs a three-branch parallel architecture handling semantics, details, and boundaries, achieving an optimal trade-off between accuracy and inference speed. Similarly, RefineMask [56] progressively optimizes instance boundaries through fine-grained feature propagation, while Mask Transfiner [57] exploits Transformers for pixel-level contour correction in boundary regions.

In the realm of multi-scale feature fusion, FaPN [58] enhances cross-scale information transfer in pyramid networks via feature alignment strategies. CARAFE [59] introduces a content-aware feature reassembly operator that adaptively generates upsampling kernels based on semantic content. DySample [60] proposes a dynamic upsampling strategy, explicitly avoiding the blurring artifacts caused by fixed interpolation. In remote sensing scenarios, FactSeg [61] decouples the learning of foreground semantics and boundary geometry via a foreground activation-driven framework. BANet [62] devises a boundary-aware network for building extraction, improving contour completeness through explicit supervision. UNetFormer [63] synergizes CNNs and Transformers, balancing global context with local details for remote sensing segmentation.

A fundamental limitation shared by these approaches is their exclusive reliance on spatial-domain features for boundary acquisition, neglecting the explicit modeling of high-frequency characteristics from a spectral perspective. When spatial signals degrade due to the smoothing effects of deep networks, subsequent boundary enhancement lacks reliable information sources. In response, this paper proposes to directly extract high-frequency boundary responses from the frequency domain, combining them with edge

confidence-guided refinement. This provides a novel pathway beyond the spatial domain for boundary-aware segmentation.

3. Method

3.1. Overall Framework

The overall architecture of the proposed Geometric Context and Frequency Domain Fusion Network (GCF-Net) is illustrated in Figure 1. The encoder extracts high-level semantic features from input images. By using the Pyramid Lightweight Fusion (PGF) blocks as depicted in Figure 1, it enhances feature representation while reducing spatial resolution, thereby enabling fine-grained information capture and improving suitability for landslide remote sensing segmentation. Specifically, we assume the input image is $F_{\text{in}} \in \mathbb{R}^{H \times W \times C}$, where H is the height of the image, W is the width, and C is the number of channels. After being processed by the encoder backbone, the output feature map is $F_e \in \mathbb{R}^{H' \times W' \times C'}$. This process is formulated as follows:

$$F_e = \text{Encoder}(F_{\text{in}}) \tag{1}$$

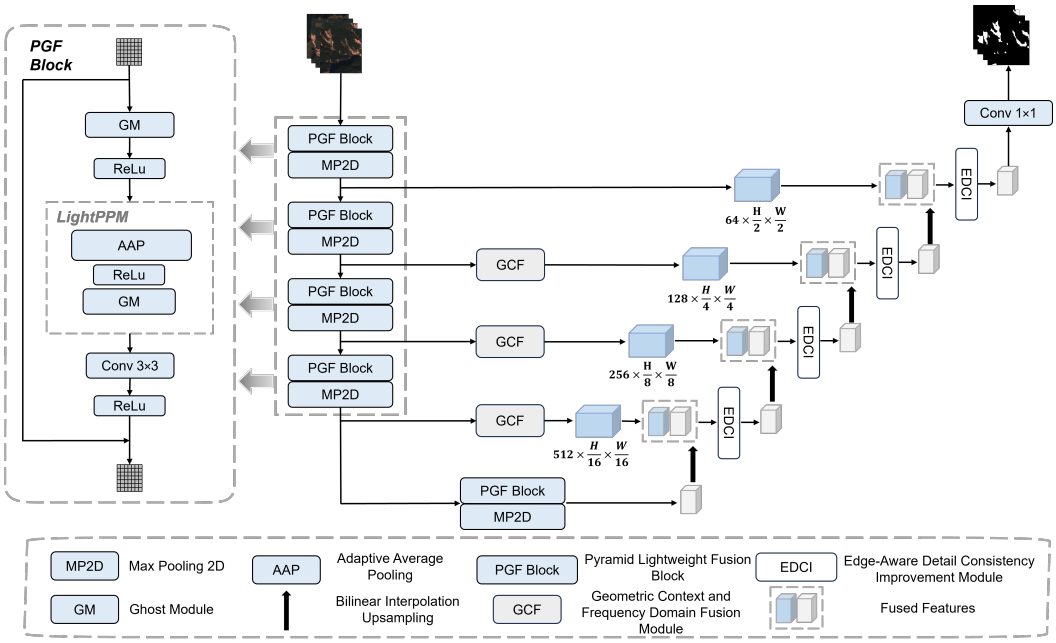


Figure 1. Overall architecture of the proposed GCF-Net. The network adopts a U-shaped encoder–decoder structure, which consists of PGF Block, GCF, and EDCI.

The decoder backbone extracts discriminative frequency-domain and contextual fusion features for classification. It consists of two main components: the Geometric Context and Frequency Domain Fusion (GCF) module as illustrated in Figure 1, which is integrated into the skip connections to extract high-frequency geometric details and multi-scale semantic contextual features, and the Edge-aware Detail Consistency Improvement (EDCI) module as depicted in Figure 1, which refines boundaries and restores semantic consistency. This process is formulated as follows:

$$F_{\text{up}} = \text{Upsample}(F_{\text{bridge}}) \tag{2}$$

$$F_{\text{out}} = \text{Decoder}(F_{\text{up}}, F_{\text{bridge}}) \tag{3}$$

$$Y = \text{Conv}_{1\times1}(F_{\text{out}}) \tag{4}$$

where $F_{\text{bridge}} \in \mathbb{R}^{H' \times W' \times C'}$ indicates the features output from the skip connections enhanced by GCF modules. These upsampled features are then fused with the encoder's feature maps in the decoder to produce the final output feature map F_{out}. Finally, the segmentation result Y is generated via the decoder's final convolutional layer.

3.2. Pyramid Lightweight Fusion Block

Accurate delineation in landslide image segmentation necessitates the extraction of features that are both structure-sensitive and detail-preserving, a task significantly complicated by complex terrain conditions. However, conventional convolutional operations frequently generate redundant features and struggle to capture fine-grained local structures, thereby constraining segmentation performance. To address these limitations, we propose the PGF block, which is designed to enhance the effectiveness and adaptability of feature representations.

The PGF block serves as a preliminary feature processing module for the subsequent GCF and EDCI modules. By employing a pyramid pooling structure combined with lightweight Ghost modules, it efficiently captures contextual information across multiple receptive fields without heavy computational cost. This multi-scale design is particularly critical for handling small-scale landslides. Since minute targets often vanish during consecutive down-sampling operations, the PGF module utilizes a residual connection mechanism to fuse the extracted global context with the original input features. This ensures that the fine-grained visual signatures of small targets are preserved and distinct within the feature space. This pre-processing not only reduces feature redundancy but also provides a stronger foundation for downstream modules to model geometric context and enforce boundary consistency. By providing enriched and structured feature maps, the PGF block ensures that the GCF and EDCI modules can effectively focus on geometric context modeling and boundary consistency refinement. The structure of the PGF block is illustrated on the left side of Figure 1, and its computational flow is formulated as follows:

$$\mathbf{F}_{\text{reduce}} = \text{GhostModule}(\mathbf{F}_{\text{in}}, C_{\text{mid}}), \tag{5}$$

$$\mathbf{F}_{\text{pool},s} = \text{AvgPool}(\mathbf{F}_{\text{reduce}}, s), \quad s = 1, 2, \ldots, S, \tag{6}$$

$$\mathbf{F}_{\text{LightPPM}} = \sum_{s=1}^{S} \text{GhostModule}(\mathbf{F}_{\text{pool},s}, C_{\text{out}}), \tag{7}$$

$$\mathbf{F}_{\text{out}} = \mathbf{F}_{\text{in}} + \text{Conv2d}(\mathbf{F}_{\text{LightPPM}}, C_{\text{out}}). \tag{8}$$

where $F_{\text{in}} \in \mathbb{R}^{H \times W \times C}$ implies the input feature map; C_{mid} and C_{out} signify the number of intermediate channels in the ghost convolution module and the number of output channels from the convolution operation, respectively; H, W, and C represent the image height, image width, and number of channels, respectively; F_{reduce} indicates feature map processed by the ghost module; $\text{AvgPool}(\cdot, s)$ denotes an average pooling operation with kernel size s, producing $\mathbf{F}_{\text{pool},s}$, which captures contextual features at multiple scales $s = 1, 2, \ldots, S$. Pyramidal pooling integrates multi-scale features through average pooling at each scale, yielding the final output F_{LightPPM}. F_{out} represents the final output feature map.

3.3. Geometric Context and Frequency Domain Fusion Module

Precise boundary delineation and robust feature representation in low-texture regions are pivotal for the performance of remote sensing landslide segmentation models. This requirement dictates that feature representations must satisfy two primary criteria. First, features must possess sufficient spatial resolution and structural sensitivity to preserve boundary continuity and capture local detail variations. Therefore, distinct from existing methods, our module is designed to be explicitly responsive to high-frequency information.

Second, features must incorporate strong contextual dependencies and discriminative power to accurately distinguish landslide boundaries from the background. To meet these demands, a novel module is required to adaptively integrate local high-frequency details with global contextual information. Diverging from prevailing methods based on global self-attention or static convolutions, this module achieves precise geometric modeling and discriminative feature extraction. Crucially, it maintains linear computational complexity through the synergistic integration of frequency-domain enhancement and multi-scale convolutions.

The architecture of the proposed GCF module is illustrated in Figure 2, which consists of two complementary branches. Branch A is designed to enhance boundary and fine-scale structures by exploiting frequency-domain characteristics. According to spectral analysis theory, edge and texture information is predominantly encoded in high-frequency components. Given the input feature map $F_{\text{in}} \in \mathbb{R}^{C\times H\times W}$, we first project it into a latent representation X through a 1×1 convolution:

$$X = \text{Conv}_{1\times 1}(F_{\text{in}}). \tag{9}$$

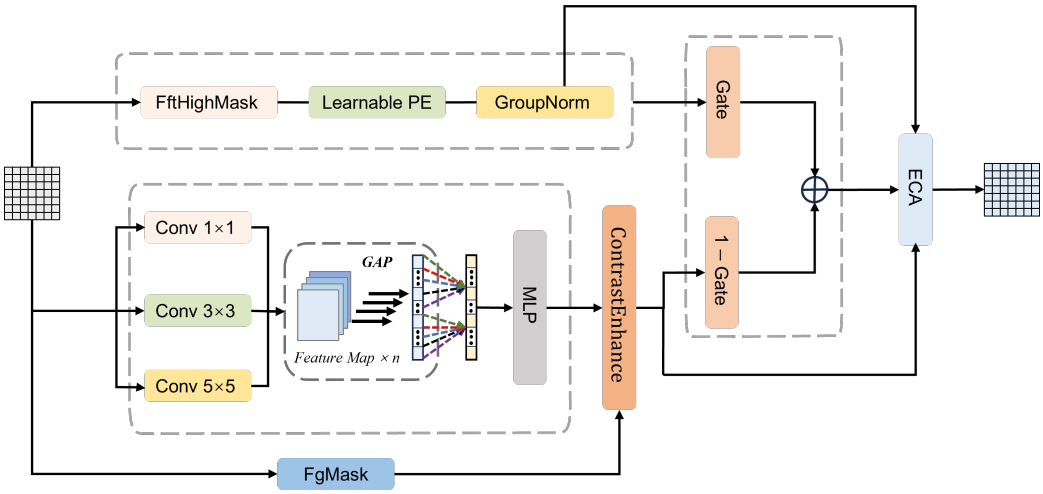


Figure 2. The architecture of the proposed GCF module.

In the first stage, we isolate high-frequency responses using spectral analysis. A two-dimensional Fast Fourier Transform (FFT) is applied to transform X into the frequency domain. Then, a static high-pass filter is employed to suppress low-frequency background information. The high-frequency spatial response H_{freq} is reconstructed via the Inverse FFT (IFFT), formulated as:

$$\mathcal{F} = \text{Shift}(\text{FFT}(X)), \quad H_{\text{freq}} = \text{Real}\Big(\text{IFFT}\Big(\text{Shift}^{-1}(\mathcal{F} \odot M_{\text{ideal}})\Big)\Big), \tag{10}$$

where $\mathcal{F}$ denotes the shifted frequency spectrum, and M_{ideal} is a binary high-pass mask constructed to suppress low-frequency components located at the spectral center, retaining the high-frequency structural information explicitly.

Subsequently, to adaptively modulate these high-frequency features based on spatial content, we incorporate a Spatial Dynamic Masking mechanism. Unlike static filtering, this module generates a learnable mask $M_{\text{sp}} \in [0,1]^{1\times H\times W}$ predicted from the input features using a lightweight convolutional network:

$$M_{\text{sp}} = \sigma\big(\text{Conv}_{1\times 1}\big(\delta(\text{Conv}_{1\times 1}(X))\big)\big), \tag{11}$$

where σ and δ denote Sigmoid and ReLU activation functions, respectively. The spatially modulated feature X_{sp} is then computed by fusing the high-frequency component with the original feature under the guidance of this mask:

$$X_{sp} = M_{sp} \odot H_{freq} + (1 - M_{sp}) \odot X, \tag{12}$$

where $\odot$ denotes element-wise multiplication. This operation ensures that high-frequency details are selectively enhanced in texture-rich regions. Crucially, unlike prior methods that indiscriminately amplify high-frequency components, this mechanism introduces spatial selectivity. Through the element-wise weighting of H_{freq} by the learnable mask M_{sp}, the module transforms global enhancement into content-aware modulation. This design allows the model to focus on enhancing structural details in salient regions while attenuating high-frequency responses in less informative areas, thus effectively mitigating the potential for noise amplification.

Following the spatial modulation, channel-wise regulation and structural encoding are applied to ensure robust feature representation. A channel-wise adaptive weighting factor $\eta \in [0, 1]^C$ is computed via Global Average Pooling (GAP) and a Multilayer Perceptron (MLP) to balance global semantic consistency. The channel-refined feature X_{fuse} is obtained by:

$$\eta = \text{MLP}(\text{GAP}(X)), \qquad X_{fuse} = \eta \odot X_{sp} + (1 - \eta) \odot X. \tag{13}$$

Finally, a Learnable Positional Encoding (LearnablePE) is injected to compensate for spatial information loss in the frequency domain, followed by Group Normalization (GroupNorm) to stabilize the distribution. The formulation is as follows:

$$F_{pe} = \text{LearnablePE}(X_{fuse}) \tag{14}$$

$$F_{out} = \text{GroupNorm}(F_{pe}) \tag{15}$$

where F_{pe} denotes the position-encoded feature, and F_{out} represents the final normalized output feature.

Branch B focuses on multi-scale contextual information and foreground enhancement. To mitigate information redundancy from conventional multi-scale concatenation, we designed a multi-scale adaptive weighting module that employs global average pooling and multilayer perceptron to compute attention weights for three scales. This mechanism allows the network to adaptively modulate the contribution of each scale based on the global context. Specifically, by assigning higher weights to larger kernels for expansive targets and emphasizing smaller kernels for fine-grained details, the module effectively accommodates both large-scale and small-scale landslides. Subsequently, a contrast enhancement unit performs explicit foreground–background contrast modeling. This strategy improves landslide discrimination in low-contrast regions while mitigating the background noise. The forward propagation is formulated as follows:

$$F_1 = \text{Conv}_{1\times1}(F_{in}), \quad F_2 = \text{Conv}_{3\times3}(F_{in}), \quad F_3 = \text{Conv}_{5\times5}(F_{in}) \tag{16}$$

$$F_{fuse} = \text{MLP}(\text{GAP}(F_1, F_2, F_3)) \tag{17}$$

$$F_{context} = \text{ContrastEnhance}(F_{fuse}, \text{FgMask}(F_{in})) \tag{18}$$

where F_1, F_2, and F_3 represent the feature maps obtained through different convolutional kernels; GAP and MLP denote global average pooling and multi-layer perceptron for adaptive weight prediction; ContrastEnhance indicates contrast enhancement; and F_{fuse},

F_{context}, and FgMask refer to the fused feature map, the enhanced feature map, and the foreground encoding, respectively.

In the feature fusion stage, a gating mechanism (Gate) adaptively integrates high-frequency features from Branch A with the multi-scale Branch B. It evaluates the significance of each feature position-wise, regulating information flow while suppressing redundancy and retaining key features. This concept is analogous to multi-layer feature fusion methods, such as gated fully fusion. Finally, the fused features undergo channel recalibration through the efficient channel attention (ECA) module [64]. The forward propagation is formulated as follows:

$$\text{Fused} = \text{Gate} \cdot F_{\text{out}} + (1 - \text{Gate}) \cdot F_{\text{context}} \tag{19}$$

$$F_{\text{fin}} = \text{ECA}(\text{Fused} + F_{\text{out}} + F_{\text{context}}) \tag{20}$$

where Fused signifies the feature map after gating-based fusion; Gate represents the gating weight for Branch A; $1 - \text{Gate}$ refers to the gating weight for Branch B; ECA indicates the standard ECA module; and F_{fin} signifies the final output feature map.

3.4. Edge-Aware Detail Consistency Improvement Module

Complementing the GCF module, the EDCI module further refines the enhanced features by focusing on feature consistency and boundary constraints. Building upon the boundary-sensitive representations and high-frequency detail enhancements provided by the GCF module, the EDCI module is designed to further improve spatial continuity in boundary regions and local prediction consistency, particularly for fine-grained contours and small-scale targets. By explicitly modeling boundary confidence and adaptively adjusting feature smoothing and detail preservation near edges, the EDCI module jointly enforces contour continuity and local consistency. As illustrated in Figure 3, it reinforces boundary integrity and the structural representation of small target regions while complementing the global semantic and high-frequency enhancements provided by the GCF module.

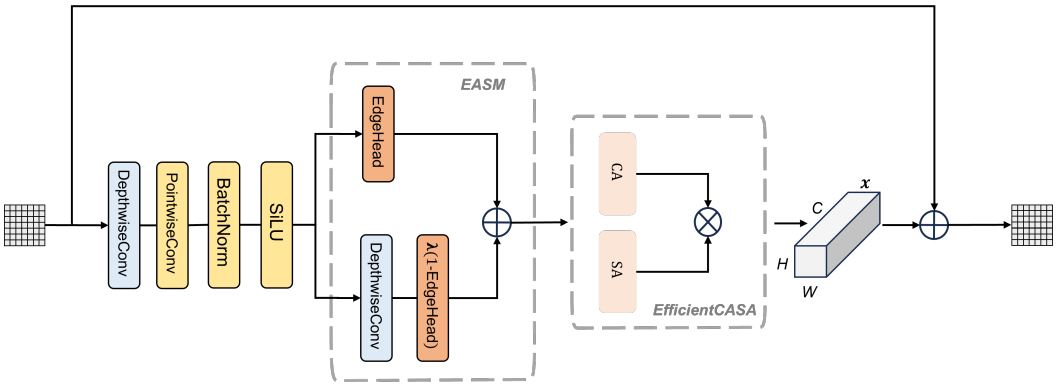


Figure 3. The architecture of the proposed EDCI module.

The key innovation of the EDCI module lies in Edge-Adaptive Smoothing Module (EASM), which integrates a boundary-aware smoothing mechanism with a learnable smoothing coefficient λ for boundary-adaptive optimization. Efficient channel features are utilized by the boundary confidence prediction head to produce a per-pixel boundary confidence map $E(x)$. A learnable smoothing coefficient λ is employed to adaptively fuse features, balancing noise suppression and boundary preservation. This approach maintains original features in high-confidence boundary regions for contour continuity while enhancing local consistency in low-confidence areas through adaptive smoothing. To enhance sensitivity to boundaries and small targets, the module employs channel and spatial attention mechanisms, emphasizing critical channels and regions to strengthen boundary feature representation. A residual connection then integrates the original and

enhanced features, preserving low-level details while maintaining high-level semantics. Together, the GCF and EDCI modules provide a complementary framework that enhances both global semantic representations and local boundary fidelity for accurate segmentation in complex terrains.The EDCI module is formulated as follows:

$$\begin{aligned} F_{\text{act}} &= \text{SiLU}(\text{BatchNorm}(\text{PointwiseConv} \\ &\quad (\text{DepthwiseConv}(F_{\text{in}})))) \end{aligned} \tag{21}$$

$$\begin{aligned} F_{\text{eas}} &= \text{EASM}(F_{\text{act}}) \\ &= E \odot F_{\text{act}} + \lambda \cdot (1 - E) \odot F_{\text{smooth}} \end{aligned} \tag{22}$$

$$F_{\text{att}} = \text{EfficientCASA}(F_{\text{eas}}) \tag{23}$$

$$F_{\text{out}} = \text{SiLU}(F_{\text{in}}^{\text{res}} + F_{\text{att}}) \tag{24}$$

where $F_{\text{in}} \in \mathbb{R}^{H \times W \times C}$ represents the input feature map; H, W, and C indicate the feature map height, width, and number of channels, respectively; DepthwiseConv and PointwiseConv indicate the depthwise separable convolution and pointwise convolution operations, respectively; $E = \text{EdgeHead}(F_{\text{act}})$ and $F_{\text{smooth}} = \text{DepthwiseConv}(F_{\text{act}}, C_{\text{out}})$; $\odot$ denotes element-wise multiplication; λ implies the learnable smoothing coefficient; F_{eas} refers to the weighted feature map obtained through this approach; EfficientCASA signifies the efficient channel-spatial attention module; F_{att} represents the feature map processed by this module; $F_{\text{in}}^{\text{res}}$ refers to the residual; and F_{out} indicates the final output feature map after fusion.

3.5. Joint Loss Function

A joint loss function integrating Focal [65], Dice [66], and Edge losses is employed to simultaneously optimize pixel-level classification accuracy, mask overlap, and boundary prediction accuracy, respectively. The combined loss function is defined as follows:

$$L_{\text{total}} = L_{\text{Focal}} + L_{\text{Dice}} + \lambda L_{\text{Edge}} \quad (\lambda > 0) \tag{25}$$

where λ denotes the edge loss weight coefficient that regulates the contribution of edge supervision to the total loss, and L_{Focal} refers to the focal loss, which is calculated as follows:

$$\begin{aligned} L_{\text{Focal}} = -\frac{1}{N} \sum_{i=1}^{N} \Big[& \alpha (1 - S_i^{\text{pred}})^{\gamma} S_i^{\text{GT}} \log(S_i^{\text{pred}}) \\ & + (1 - \alpha)(S_i^{\text{pred}})^{\gamma} (1 - S_i^{\text{GT}}) \log(1 - S_i^{\text{pred}}) \Big] \end{aligned} \tag{26}$$

where S_i^{pred} indicates the predicted pixel probability; S_i^{GT} represents the ground truth label; α and γ refer to focal loss hyperparameters; N indicates the total number of pixels; and L_{Dice} represents the dice loss function, which quantifies the overlap between the predicted and the ground truth masks. It is calculated as follows:

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^{N} S_i^{\text{pred}} S_i^{\text{GT}} + \epsilon}{\sum_{i=1}^{N} (S_i^{\text{pred}})^2 + \sum_{i=1}^{N} (S_i^{\text{GT}})^2 + \epsilon} \tag{27}$$

where S_i^{pred} indicates the predicted pixel probability; S_i^{GT} represents the ground truth label; ϵ represents a smoothing term to prevent division by zero. For multi-channel masks, the loss is computed per channel separately and then averaged.

Edge loss improves boundary precision by comparing the gradients of the predicted and ground-truth edges, extracted using the Sobel operator, and supervised with binary cross-entropy. It is calculated as follows:

$$L_{\text{Edge}} = \frac{1}{C}\sum_{c=1}^{C} \text{BCE}\left(\hat{E}_P^{(c)}, \hat{E}_G^{(c)}\right) \tag{28}$$

$$\hat{E}_P^{(c)} = \sigma\left(\sqrt{(P_x^{(c)})^2 + (P_y^{(c)})^2 + \epsilon}\right), \tag{29}$$

$$\hat{E}_G^{(c)} = \frac{\sqrt{(G_x^{(c)})^2 + (G_y^{(c)})^2 + \epsilon}}{\max(E_G^{(c)}) + \epsilon} \tag{30}$$

where $P_x^{(c)}$, $P_y^{(c)}$, $G_x^{(c)}$, and $G_y^{(c)}$ represent the Sobel gradient convolution results of the predicted and ground truth masks in the x and y directions, respectively, and σ denotes the sigmoid activation function.

This joint optimization strategy enables the model to concurrently minimize pixel-level classification loss, region-level overlap loss, and boundary precision loss, offering a holistic evaluation of segmentation performance. It alleviates class imbalance, enhances detection of small or indistinct targets, and strengthens overall model generalization. Furthermore, it reduces the gap between the predicted segmentation P^{pred} and the ground truth segmentation G and aligns the predicted edges $\hat{E}_P$ and the ground truth edges $\hat{E}_G$, thereby ensuring precise and well-defined contours. This joint optimization strategy enables precise boundary delineation without compromising overall segmentation accuracy.

4. Experiments

4.1. Data Description

To rigorously assess the proposed model, three landslide datasets were employed: the Bijie City (Guizhou Province) dataset [67], the INSAR landslide dataset (2022) [68], and the Large-scale Multi-source High-resolution Landslide Dataset (LMHLD) [69]. Due to the limited size of the Bijie dataset (770 raw images), it was combined with the INSAR landslide dataset to create a Hybrid dataset capturing diverse landslide morphologies. The final experiments utilized this hybrid dataset and the LMHLD dataset to evaluate the model's generalization across varying landslide types and scenarios. Specifically, the datasets encompass a wide range of landslide types and scenarios, as summarized in Table 1.

Table 1. Summary of landslide types and scenarios in the datasets.

Category	Type/Scenario	Description
Morphological Types	Large-scale continuous	Single large-scale slope failure with clear and complete boundaries.
	Small-scale scattered	Multiple fragmented landslides sparsely distributed across the slope.
	Elongated strip-shaped	Narrow landslides extending along channels or steep hillslopes.
Environmental Scenarios	Forested slope	Landslides surrounded by dense vegetation cover.
	Roadside	Landslides along transportation routes and artificial cut slopes.

4.1.1. Hybrid Dataset

The experimental dataset, referred to as the Hybrid Dataset, integrates data from two distinct sources: the Bijie City dataset and the INSAR landslide dataset. The Bijie dataset, captured by TripleSat satellites from May to August 2018, consists of 770 high-resolution (0.8 m) raw images characterized by pronounced topographic relief and abundant rainfall. The INSAR dataset encompasses 3799 raw images derived from diverse global geomorphologies, primarily acquired by Sentinel-2 satellites with resolutions ranging from 10 m to 60 m. In total, the Hybrid Dataset comprises 4569 raw images. Representative samples, corresponding ground truth maps, and visualization results are illustrated in Figure 4.

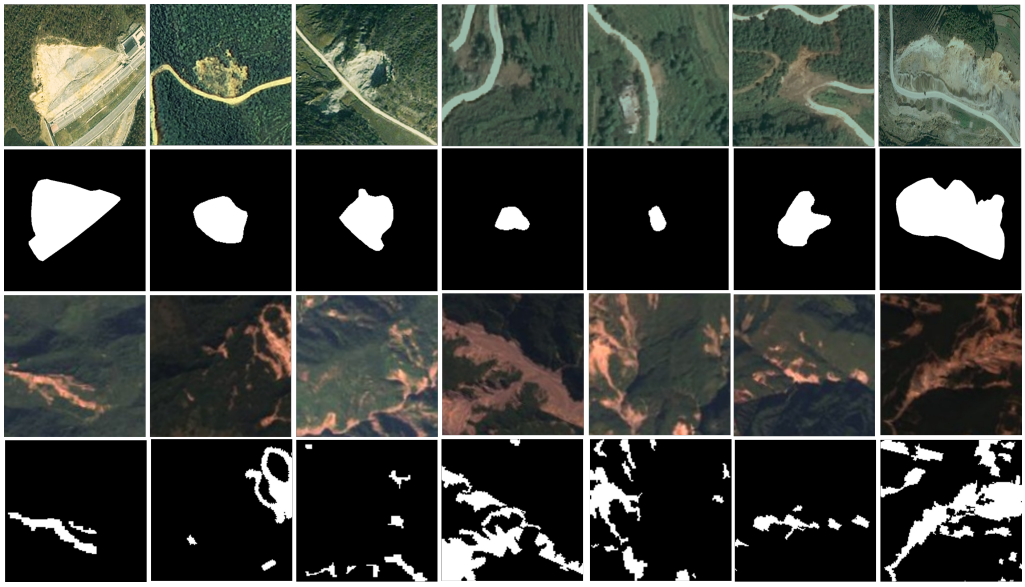


Figure 4. Original image and label image of the Hybrid dataset.

4.1.2. LMHLD Dataset

Developed by the School of Future Technology at China University of Geosciences (Wuhan), this dataset integrates remote sensing imagery from five distinct satellite sensors across seven representative global landslide regions. The dataset encompasses landslide events from Wenchuan, China (2008); Rio de Janeiro, Brazil (2011); Gorkha, Nepal (2015); Jiuzhaigou, China (2015); Taiwan, China (2018); Hokkaido, Japan (2018); and Emilia-Romagna, Italy (2023). From this dataset, a total of 6776 raw images characterized by varying dimensions were selected to capture landslides across different scales. Representative samples of these raw images, along with their corresponding ground truth maps and relevant metadata, are illustrated in Figure 5.

4.2. Experimental Setup

4.2.1. Dataset Configuration

The proposed model accepts fixed-size inputs of 256 × 256 pixels. Since raw images in the datasets have varying dimensions, all images are resized while preserving their aspect ratios and padded with gray pixels to achieve the uniform target size of 256 × 256 during the data loading phase. Additionally, remote sensing images were subjected to channel processing to mitigate the impact of heterogeneity among different data sources. During training, data augmentation techniques including random scaling, random horizontal flipping, and color jittering are applied to enhance sample diversity and improve model generalization. The dataset was randomly split into training, validation, and test sets with a ratio of 7:2:1, and data augmentation was applied to the training set only.

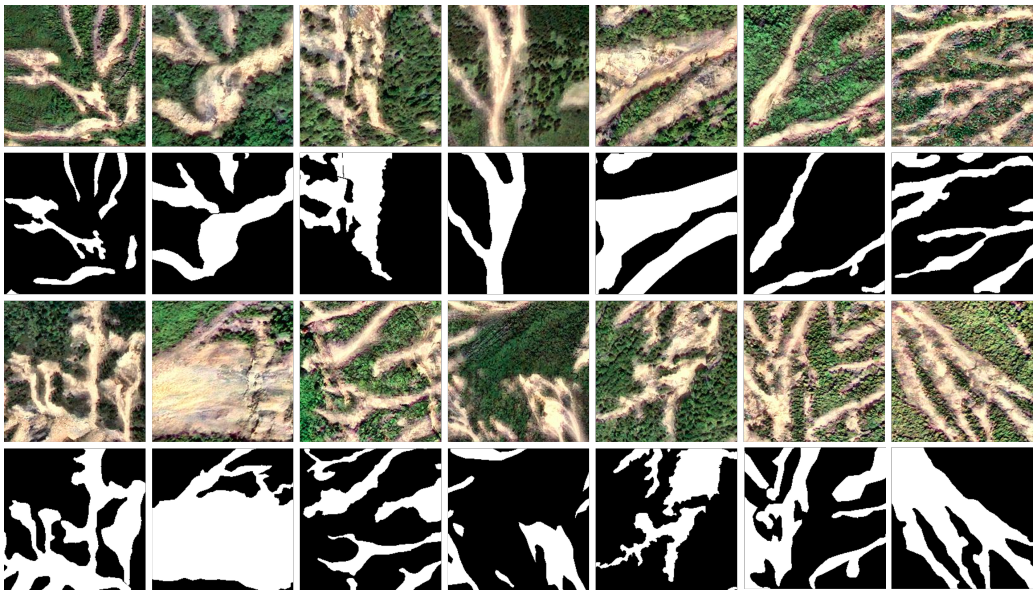


Figure 5. Original image and label image of the LMHLD dataset.

4.2.2. Training Details

All experiments were implemented using PyTorch 2.0 and Python 3.10, running on an NVIDIA GeForce RTX 4090 GPU (24 GB memory) (Nvidia, Santa Clara, CA, USA) with a 12-core Intel Xeon Platinum 8375C CPU (2.90 GHz) (Intel, Santa Clara, CA, USA) under Windows 10. To ensure reproducibility, we summarize the key training hyperparameters in Table 2.

Table 2. Summary of training hyperparameters.

Hyperparameter	Value
Input size	256×256
Batch size	8
Optimizer	Adam
Initial learning rate	1×10^{-4}
Minimum learning rate	1×10^{-6}
β_1, β_2	0.9, 0.999
Weight decay	0
Learning rate scheduler	Cosine annealing
Training epochs	100
Early stopping patience	10 epochs

4.2.3. Evaluation Metrics

Six widely used evaluation metrics in landslide extraction tasks were adopted, including Overall Accuracy (OA), Average Accuracy (AA), Kappa coefficient (Kappa), F1-score (F1), Recall (R), and mean Intersection over Union (mIoU). Higher values of these metrics indicate better model performance.The metrics are defined as:

$$\mathrm{OA} = \frac{TP + TN}{TP + FP + FN + TN} \tag{31}$$

$$\mathrm{AA} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{PA}_i \tag{32}$$

$$\kappa = \frac{\mathrm{OA} - P_e}{1 - P_e} \tag{33}$$

$$\text{F1} = 2 \times \frac{P \times R}{P + R} \tag{34}$$

$$R = \frac{TP}{TP + FN} \tag{35}$$

$$\text{mIoU} = \frac{1}{n} \sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{36}$$

In these equations, TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. P denotes precision ($P = \frac{TP}{TP+FP}$), and n is the number of classes. PA_i indicates the pixel accuracy for class i, and P_e is the expected agreement by chance.

Furthermore, to rigorously evaluate the boundary quality and the model's performance on the minority landslide class, we incorporate three boundary-sensitive and foreground-aware metrics: Positive-Class IoU (P-IoU), Boundary F1-score (B-F1), and 95% Hausdorff Distance (HD95). P-IoU assesses the overlap specifically for the landslide category, avoiding the dominance of background pixels. B-F1 evaluates the precision and recall of the predicted contours within a specific boundary region, while HD95 quantifies the shape error by calculating the distance between the predicted and ground truth boundaries. These metrics are defined as:

$$\text{P-IoU} = \frac{TP}{TP + FP + FN} \tag{37}$$

$$\text{B-F1} = \frac{2 \times P_b \times R_b}{P_b + R_b} \tag{38}$$

$$\text{HD95} = \max\left(K_{95\%}\{\min_{y \in Y} ||x - y||\}_{x \in X},\ K_{95\%}\{\min_{x \in X} ||y - x||\}_{y \in Y} \right) \tag{39}$$

where P_b and R_b denote the boundary precision and boundary recall, respectively, calculated based on pixel matching within a specified distance threshold. In Equation (39), X and Y represent the set of ground truth and predicted boundary points, respectively, and the operator $K_{95\%}$ indicates the 95th percentile of the Euclidean distances.

4.3. Comparative Experiments

To rigorously evaluate the proposed model, we conducted a comprehensive benchmarking study against eight representative segmentation methods. The selected comparators, including U-Net [6], SwinUnet [11], TransUNet [70], DAE-Former [18], MERIT [71], G-CASCADE [17], RS3Mamba [12], and SegMAN [13], were chosen to ensure architectural diversity, temporal currency, and methodological relevance. This selection spans CNN, Transformer, and Mamba architectures, effectively capturing the evolution from the classical U-Net to the recent SegMAN. Notably, G-CASCADE is included for its explicit focus on boundary enhancement, while the remaining baselines leverage advanced multi-scale feature extraction and attention mechanisms capable of capturing complex semantic dependencies. To ensure a fair comparison, all models were trained from scratch on our datasets under uniform experimental protocols. While an exhaustive comparison with every existing method is infeasible, this diverse set provides a robust baseline for demonstrating the effectiveness of our approach. Furthermore, we adopted specialized evaluation metrics to quantitatively assess the model's performance in boundary-aware segmentation.

4.3.1. Comparison on the Hybrid Dataset

Table 3 presents a comparison between the Geometric Context and Frequency Domain Fusion Network (GCF-Net) and representative classification methods on the Hybrid dataset. GCF-Net consistently outperforms existing approaches across all comprehensive metrics, achieving an OA of 96.42%, AA of 90.28%, and a Kappa coefficient of 83.90%. Notably, compared to the second-best method, SegMAN, our model demonstrates substantial improvements, increasing the F1-score and mIoU by 4.34% and 4.21%, respectively. Regarding Recall, GCF-Net maintains a competitive level comparable to SegMAN but distinguishes itself by achieving a superior balance without sacrificing precision. This advantage is further evidenced by the remarkable mIoU gain, indicating that our dual-domain fusion strategy effectively captures boundary details and reduces false positives under similar detection capabilities, resulting in more topologically accurate predictions.

Table 3. Comparison of Experimental Results of Nine Models on the Hybrid Dataset.

Year	Model	OA (%)	AA (%)	Kappa (%)	F1 (%)	Recall (%)	mIoU (%)
2015	U-Net	91.02	85.50	61.66	66.71	78.32	70.09
2022	SwinUnet	90.37	88.21	63.74	67.08	83.41	69.89
2021	TransUNet	91.57	89.86	65.79	70.49	81.64	72.53
2023	DAE-Former	93.12	87.53	70.54	74.42	84.17	75.81
2024	MERIT	94.08	87.37	71.98	75.34	78.64	76.97
2024	G-CASCADE	94.80	87.43	75.98	78.93	84.86	79.71
2024	RS3Mamba	95.07	88.48	77.12	79.96	84.90	80.52
2025	SegMAN	95.29	88.86	77.93	80.60	85.12	81.14
2025	**GCF-Net (Ours)**	**96.42**	**90.28**	**83.90**	**84.94**	**85.20**	**85.35**

4.3.2. Comparison on the LMHLD Dataset

As presented in Table 4, the proposed GCF-Net achieves state-of-the-art performance on the LMHLD dataset, outperforming the second-best method, SegMAN, across all metrics. Specifically, GCF-Net attains an OA of 96.71%, AA of 89.59%, and F1-score of 85.85%. A critical observation is that while the improvement in OA is 0.83%, the gain in AA is much more substantial at 3.04% compared to SegMAN. Since AA represents the average accuracy across categories, this significant boost indicates that GCF-Net effectively mitigates the class imbalance problem, drastically improving the recognition capability for the minority landslide class rather than merely overfitting to the background. Furthermore, the remarkable increase of 5.32% in F1-score demonstrates the model's superior discriminative power, ensuring high-quality segmentation with fewer false positives compared to existing transformers and CNN-based methods.

Table 4. Comparison of Experimental Results of Nine Models on the LMHLD Dataset.

Year	Model	OA (%)	AA (%)	Kappa (%)	F1 (%)	Recall (%)	mIoU (%)
2015	U-Net	94.43	82.06	69.31	69.38	66.82	73.59
2022	SwinUnet	94.62	85.82	71.37	72.47	74.97	75.52
2021	TransUNet	94.74	84.81	73.46	72.28	72.56	75.47
2023	DAE-Former	95.25	86.44	77.50	76.14	76.04	77.53
2024	MERIT	95.52	84.57	78.13	77.96	78.26	78.11
2024	G-CASCADE	95.74	84.93	79.57	79.61	80.75	79.04
2024	RS3Mamba	95.63	85.71	79.13	79.50	80.12	79.29
2025	SegMAN	95.88	86.55	81.08	80.53	81.34	81.64
2025	**GCF-Net (Ours)**	**96.71**	**89.59**	**84.46**	**85.85**	**83.35**	**83.89**

Figures 6 and 7 visualize the qualitative segmentation results of all competing methods on the Hybrid and LMHLD datasets. As observed, SwinUnet and TransUNet exhibited relatively inferior performance, frequently struggling to distinguish landslide boundaries

from complex backgrounds, leading to noticeable false positives. Conversely, DAE-Former and MERIT demonstrated superior classification capabilities by effectively capturing global semantic context; however, the proposed method further minimized misclassification errors in ambiguous transition zones. By recovering fine-grained geometric details, our approach achieved the highest classification accuracy and generated the most structurally precise landslide masks compared to other state-of-the-art baselines.

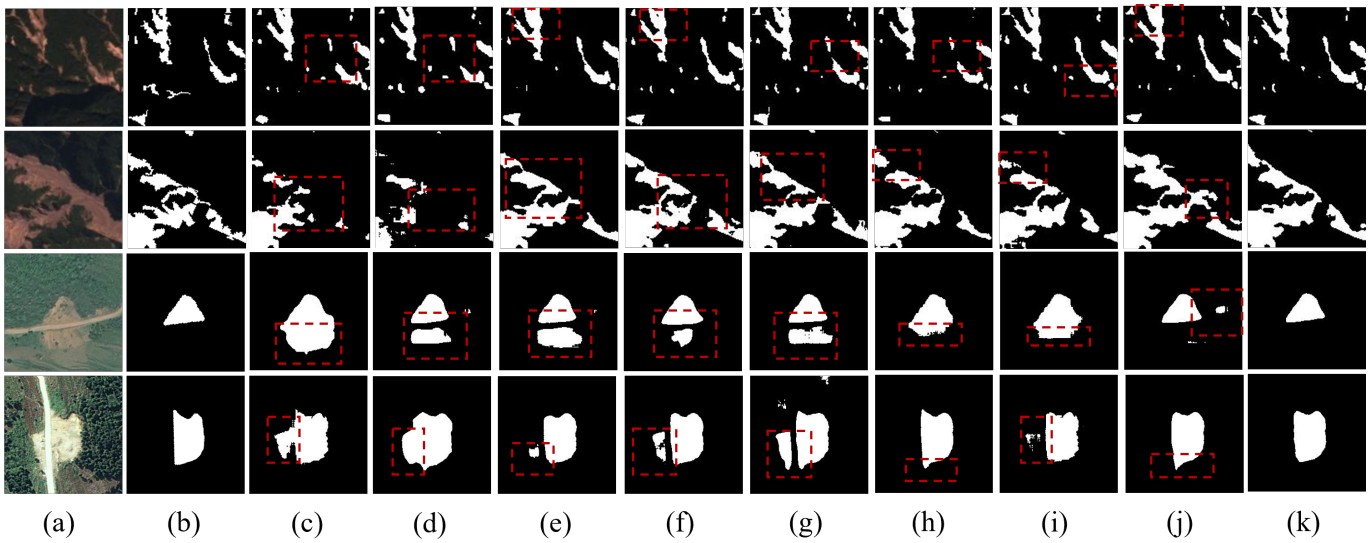


Figure 6. Extraction results of the nine models on the test set. (**a**) Hybrid dataset images, (**b**) ground truth, (**c**) U-Net, (**d**) SwinUnet, (**e**) TransUNet, (**f**) DAE-Former, (**g**) MERIT, (**h**) G-CASCADE, (**i**) RS3Mamba, (**j**) SegMAN, (**k**) GCF-Net. The red dashed boxes represent landslides extracted by different methods, which show significant differences compared to the ground truth.

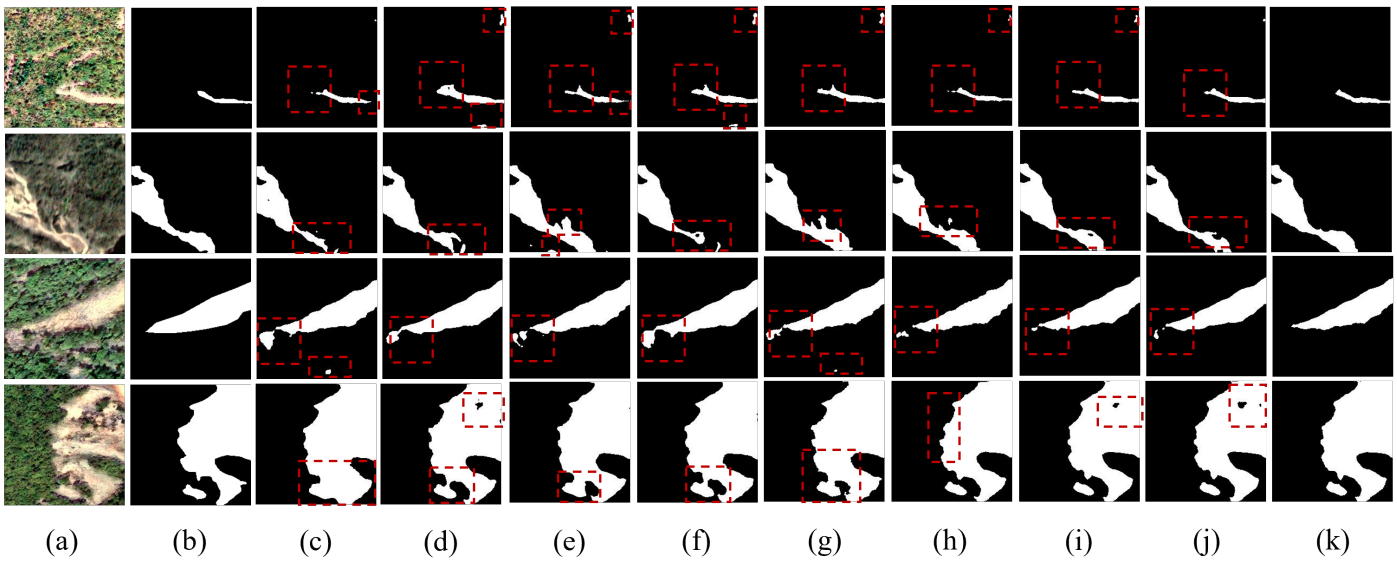


Figure 7. Extraction results of the nine models on the test set. (**a**) LMHLD dataset images, (**b**) ground truth, (**c**) U-Net, (**d**) SwinUnet, (**e**) TransUNet, (**f**) DAE-Former, (**g**) MERIT, (**h**) G-CASCADE, (**i**) RS3Mamba, (**j**) SegMAN, (**k**) GCF-Net. The red dashed boxes represent landslides extracted by different methods, which show significant differences compared to the ground truth.

4.4. Ablation Study

To validate the impact of GCF-Net's key modules on the classification performance, a series of ablation experiments were conducted on the Hybrid and LMHLD datasets. Each

core module, namely the Pyramid Lightweight Fusion (PGF) block, the Geometric Context and Frequency Domain Fusion (GCF) module, and the Edge-aware Detail Consistency Improvement (EDCI) module, were sequentially removed or substituted, and the resulting changes in OA, AA, Kappa coefficient, F1, R, and mIoU were analyzed.

Baseline Model: The baseline model employed in these ablation experiments was a symmetric U-shaped encoder-decoder network with four downsampling and four upsampling stages, each consisting of two sequential 3×3 convolutional layers with Rectified Linear Unit (ReLU) activation function. This baseline model comprised the fundamental convolutional encoder-decoder, excluding key modules such as the PGF block, GCF, and EDCI, and served as a reference for the ablation experiments. All experiments employed identical datasets, preprocessing, hyperparameters, and training protocols to ensure fairness and enable objective assessment of each module's impact on model performance.

The results are presented in Tables 5 and 6. After removing the PGF Block, GCF, and EDCI modules, the baseline model achieved OA values of 91.02% and 94.43% on the Hybrid and LMHLD datasets, respectively. When integrating the PGF Block into the baseline model, we observed that the OA values on the Hybrid and LMHLD datasets improved to 94.18% and 95.08%, respectively. This indicates that the PGF Block effectively enhanced the model's ability to capture multi-scale information through the multi-scale feature extraction mechanism, thereby improving the overall classification performance. Similarly, after integrating the GCF module into the baseline model, as detailed in Tables 5 and 6, the mIoU values on the Hybrid and LMHLD datasets increased by 9.60% and 4.07%, respectively, demonstrating that this module plays a positive role in enhancing the feature discriminability of landslide boundaries and low-texture regions. Analogously, the integration of the EDCI module into the baseline model led to increases in F1 by 12.09% and 5.38% on the Hybrid and LMHLD datasets, respectively. Notably, as shown in Tables 5 and 6, the combined integration of the PGF Block, GCF and EDCI modules yielded a more significant performance improvement, with OA values on the Hybrid and LMHLD datasets increasing by 5.40% and 2.28% compared to the baseline model. This confirms the complementarity of the three modules.

Table 5. Ablation experiment results with module combinations on Hybrid dataset.

Baseline	PGF	GCF	EDCI	OA (%)	AA (%)	Kappa (%)	F1 (%)	R (%)	mIoU (%)
✓				91.02	85.50	61.66	66.71	78.32	70.09
✓	✓			94.18	87.33	76.73	79.69	81.16	78.21
✓		✓		94.57	87.13	77.31	80.48	82.60	79.69
✓			✓	94.06	86.62	76.11	78.80	80.89	78.26
✓	✓	✓	✓	**96.42**	**90.28**	**83.90**	**84.94**	**85.20**	**85.35**

Table 6. Ablation experiment results with module combinations on LMHLD dataset.

Baseline	PGF	GCF	EDCI	OA (%)	AA (%)	Kappa (%)	F1 (%)	R (%)	mIoU (%)
✓				94.43	82.06	69.31	69.38	66.82	73.59
✓	✓			95.08	85.77	71.93	74.58	74.21	77.08
✓		✓		95.25	86.08	72.74	75.35	74.7	77.66
✓			✓	95.19	85.39	72.07	74.76	73.23	77.26
✓	✓	✓	✓	**96.71**	**89.59**	**84.46**	**85.85**	**83.35**	**83.89**

Figures 8 and 9 present the visual comparison of segmentation results under different configurations. The baseline model exhibited obvious insufficient segmentation, misclassification, and blurred boundaries. Although the single-module variants provided better recognition, their boundary delineation remained limited. In contrast, the complete GCF-Net demonstrated sharp and continuous boundaries, reducing missed detections, especially in small-scale landslide areas. These visual results fully support the quantitative metrics, validating the effectiveness of the proposed modules.

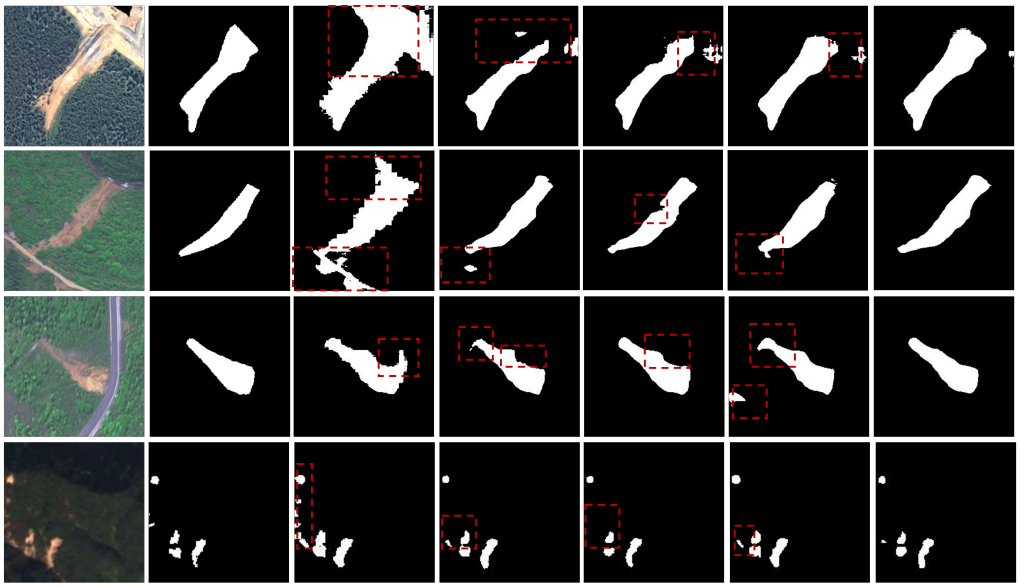


Figure 8. Visual comparison of segmentation results for different module combinations. (**a**) Hybrid dataset images, (**b**) ground truth, (**c**) Baseline, (**d**) Baseline+PGF, (**e**) Baseline+GCF, (**f**) Baseline+EDCI, (**g**) GCF-Net. Red dashed boxes highlight the regions where GCF-Net improves recognition and boundary delineation.

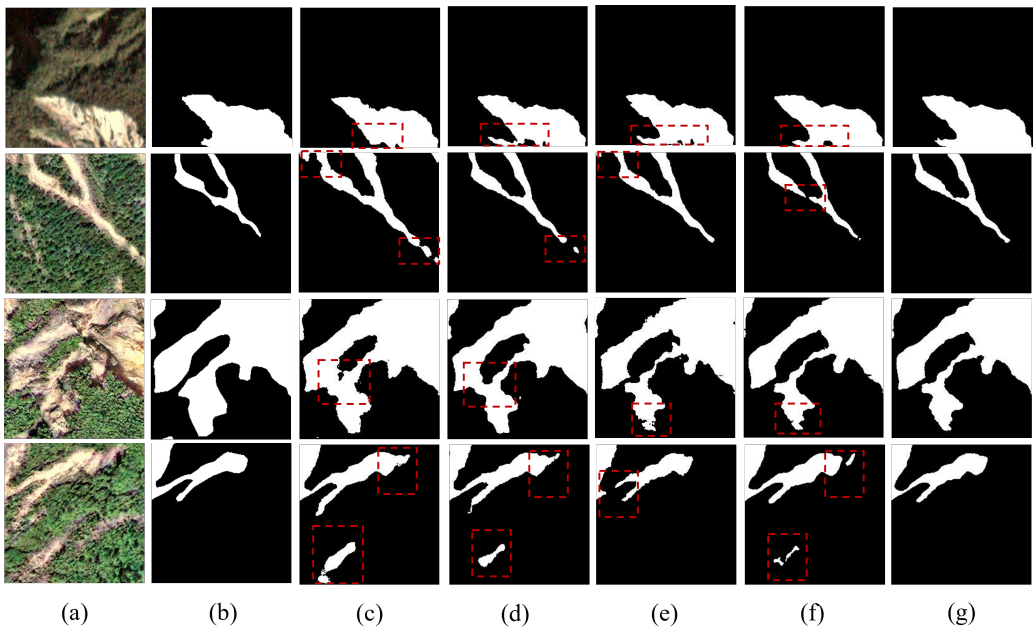


Figure 9. Visual comparison of segmentation results for different module combinations. (**a**) LMHLD dataset images, (**b**) ground truth, (**c**) Baseline, (**d**) Baseline+PGF, (**e**) Baseline+GCF, (**f**) Baseline+EDCI, (**g**) GCF-Net. Red dashed boxes highlight the regions where GCF-Net improves recognition and boundary delineation.

4.5. Analysis on Boundary Refinement Under Class Imbalance

Landslide detection often faces severe class imbalance. In our study, foreground pixels account for only 6.72% of the Hybrid dataset and 9.57% of the LMHLD dataset. Under such conditions, standard region-based metrics like OA can be misleadingly dominated by background pixels. To rigorously substantiate the boundary refinement benefits of GCF-Net, we evaluate our method using boundary-sensitive and foreground-aware metrics,

specifically Positive-Class IoU, Boundary F1-score, and 95% Hausdorff Distance, denoted as P-IoU, B-F1, and HD95, respectively.

Table 7 compares GCF-Net with the Baseline and the top-performing state-of-the-art method, SegMAN. On the Hybrid dataset, GCF-Net achieves a P-IoU of 73.82%, significantly outperforming the Baseline value of 50.05% and the SegMAN result of 67.50%. Regarding boundary precision compared to SegMAN, our method improves the B-F1 by 5.65%, rising from 57.06% to 62.71%, and effectively reduces the HD95 from 30.10 px to 25.03 px. Similar superior performance is observed on the LMHLD dataset, where GCF-Net achieves the highest B-F1 of 66.93% and the lowest HD95 of 61.24 px. These results demonstrate that by leveraging frequency domain constraints to complement spatial features, GCF-Net effectively mitigates the boundary blurring issue prevalent in existing architectures. The substantial reduction in shape error as measured by HD95 indicates that explicit high-frequency modeling enables the network to preserve fine-grained edge details that are often lost. This quantitative evidence robustly validates our claim that the proposed dual-domain strategy facilitates the refined reconstruction of landslide boundaries by suppressing background noise, even in scenarios with severe class imbalance.

Table 7. Boundary quality comparison on Hybrid and LMHLD datasets.

Dataset	Method	P-IoU (%)	B-F1 (%)	HD95 (px)
Hybrid dataset	Baseline	50.05	31.70	45.34
	SegMAN	67.50	57.06	30.10
	GCF-Net (Ours)	**73.82**	**62.71**	**25.03**
LMHLD dataset	Baseline	53.12	40.25	87.46
	SegMAN	67.41	54.70	70.29
	GCF-Net (Ours)	**75.21**	**66.93**	**61.24**

4.6. Evaluation of Generalization via Cross-Dataset Validation

Standard within-dataset random splits primarily assess in-distribution performance. To evaluate the generalization capability of GCF-Net beyond the training distribution, we conducted cross-dataset validation between the Hybrid and LMHLD datasets. This setup evaluates the model on entirely unseen data characterized by different geographic regions, landslide events, and sensor specifications. For a representative comparison, we evaluated the performance of the baseline, the top-performing state-of-the-art SegMAN method, and the proposed GCF-Net. In all tests, models were trained on the source dataset and directly evaluated on the target dataset without any fine-tuning or domain adaptation.

Table 8 summarizes the results. While all models exhibit performance decay when applied to a new dataset due to domain shift, GCF-Net demonstrates superior stability. Specifically, in the transfer test from the Hybrid dataset to the LMHLD dataset, GCF-Net achieves an OA of 81.04%, substantially outperforming the U-Net baseline by 12.33% and SegMAN by 8.50%. Furthermore, GCF-Net achieves a remarkable reduction in shape error, lowering the HD95 from 116.31 px achieved by the baseline to 43.16 px. A similar trend of performance improvement is observed when evaluating the transferability in the reverse direction from LMHLD to Hybrid. In this scenario, GCF-Net achieves the highest OA of 88.94% with a 15.30% improvement over the baseline. These results indicate that by emphasizing frequency-domain structural invariants, GCF-Net captures consistent morphological features that remain stable across heterogeneous data sources, thereby enhancing the model's generalization in unseen environments.

Table 8. Generalization results via cross-dataset validation. The Source and Target datasets represent the training and unseen evaluation data, respectively. Note that for HD95, lower values indicate better performance.

Evaluation Scenario	Method	OA (%)	AA (%)	HD95 (px)	Improv. (%)
Hybrid to LMHLD	Baseline	68.71	54.50	116.31	-
	SegMAN	72.54	63.72	73.29	+3.83
	GCF-Net (Ours)	**81.04**	**76.23**	**43.16**	**+12.33**
LMHLD to Hybrid	Baseline	73.64	43.62	122.71	-
	SegMAN	81.77	51.29	89.91	+8.13
	GCF-Net (Ours)	**88.94**	**59.01**	**71.29**	**+15.30**

5. Discussion

5.1. Complexity Analysis

The parameters, floating-point operations (FLOPs), and model size of the proposed method were compared with several other methods on the Hybrid dataset. Table 9 presents the complexity metrics of the proposed model and comparative methods, which clearly demonstrates that our model achieves a superior trade-off between segmentation accuracy and computational cost with fewer parameters and lower FLOPs. Specifically, the Geometric Context and Frequency Domain Fusion Network (GCF-Net) exhibits a smaller parameter count and significantly reduced FLOPs compared to most other networks, achieving an approximately 90.45% reduction in FLOPs relative to TransUNet. This improvement is attributed to the Ghost module, which reduces the number of parameters compared to conventional convolutions while maintaining accuracy and speed. Overall, GCF-Net achieves lower model complexity and higher computational efficiency without compromising segmentation performance, demonstrating effective architectural optimization.

Table 9. Comparison of model complexity among different methods.

Year	Method	Parameters (M)	FLOPs (G)	Model Size (MB)
2015	U-Net	24.89	225.84	94.97
2022	SwinUnet	43.44	595.11	178.34
2021	TransUNet	96.40	953.76	416.10
2023	DAE-Former	48.01	9.20	192.04
2024	MERIT	147.86	33.31	591.44
2024	G-CASCADE	24.86	3.91	99.44
2024	RS3Mamba	40.60	37.41	42.32
2025	SegMAN	29.514	23.83	**20.73**
2025	**GCF-Net (Ours)**	**9.22**	**3.43**	39.04

5.2. Analysis of GCF Layer Position Effects

To analyze the impact of the Geometric Context and Frequency Domain Fusion (GCF) modules placement at different positions on the results, GCF-Net variants with three GCF layers positioned at shallow, middle, and deep layers were evaluated on the Hybrid dataset. The specific configurations are illustrated in Figure 10. Table 10 shows that the GCF module's embedding location markedly influenced model performance. When applied to shallow-layer features, the module struggled to effectively distinguish specific features from background noise, causing foreground representations to amplify noise and misclassify non-landslide regions as foreground. Consequently, the recall rate for landslide regions declined, and overall classification consistency deteriorated. In contrast, the middle layer features, balancing detail preservation and semantic abstraction, effectively mitigated shallow-layer noise misclassification while capturing the complete structural information of the targets. This configuration improved the landslide region recall rate by 1.34% and the

F1 score by 5.76% compared to the shallow-layer features, achieving an optimal trade-off between foreground enhancement and background suppression. While deep-layer features provided rich semantic information, they suffered from significant detail loss. Their large receptive fields risked misclassifying background as foreground, thereby reducing the precision of fine-grained segmentation. In summary, integrating GCF at the middle layer optimally balances detail and semantics, underpinning the highest model performance.

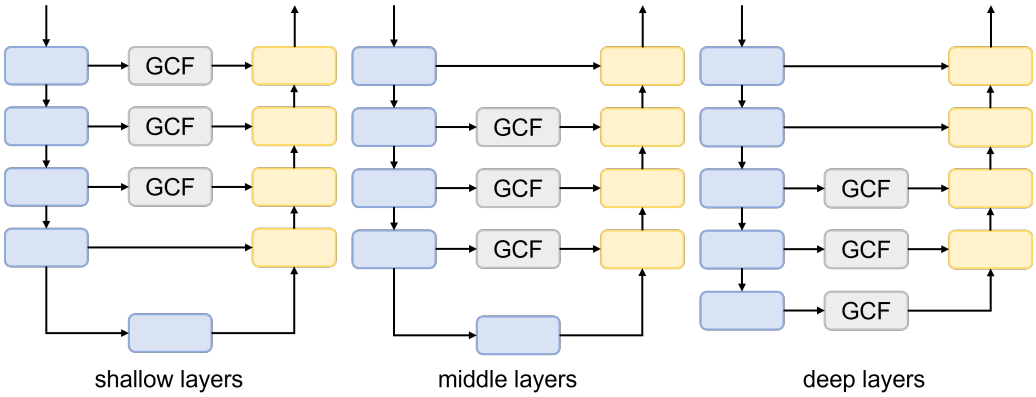


Figure 10. Structure of GCF-Net with GCF modules embedded at different positions: shallow, middle, and deep layers.

Table 10. Comparison of the effects of GCF layer positions.

Depth of GCF	OA (%)	AA (%)	Kappa (%)	F1 (%)	Recall (%)		mIoU (%)
					bk	ls	
Shallow layers	94.56	87.19	78.64	78.96	95.17	83.86	79.59
Middle layers	**96.42**	**90.28**	**83.90**	**84.94**	97.01	**85.20**	**85.35**
Deep layers	95.70	88.30	80.00	82.44	96.72	84.03	82.67

5.3. Effectiveness Analysis of EASM

To quantify the effect of the Edge-Adaptive Smoothing Module (EASM) on segmentation performance, five key segmentation metrics were selected on the Hybrid dataset to compare the performance of the fixed edge guidance strategies (thresholds 0.1, 0.3, 0.8) with the proposed EASM, and the results were plotted as a grouped bar chart, as shown in Figure 11. Compared with the fixed edge guidance strategies, the EASM achieved superior performance across all five metrics: Overall Accuracy (OA), Average Accuracy (AA), Kappa, Recall (R), and F1-score. Although the fixed edge guidance strategies could marginally improve some metrics, their overall effectiveness was limited by their inability to adapt to the spatial heterogeneity of boundary confidence. This limitation is particularly prominent in practical applications: the boundary features of landslide terrains exhibit significant variability, and fixed parameters often fail to adapt to the diverse gradients of boundary features. In contrast, the proposed EASM adaptively modulates the edge enhancement intensity via a learnable smoothing coefficient, prioritizes high-frequency regions and small target regions, and effectively improves the restoration of edge details, while ensuring robust adaptability to varying terrain conditions. The results of the grouped bar chart with trend lines also clearly demonstrate that the EASM exhibits superior performance across all metrics.

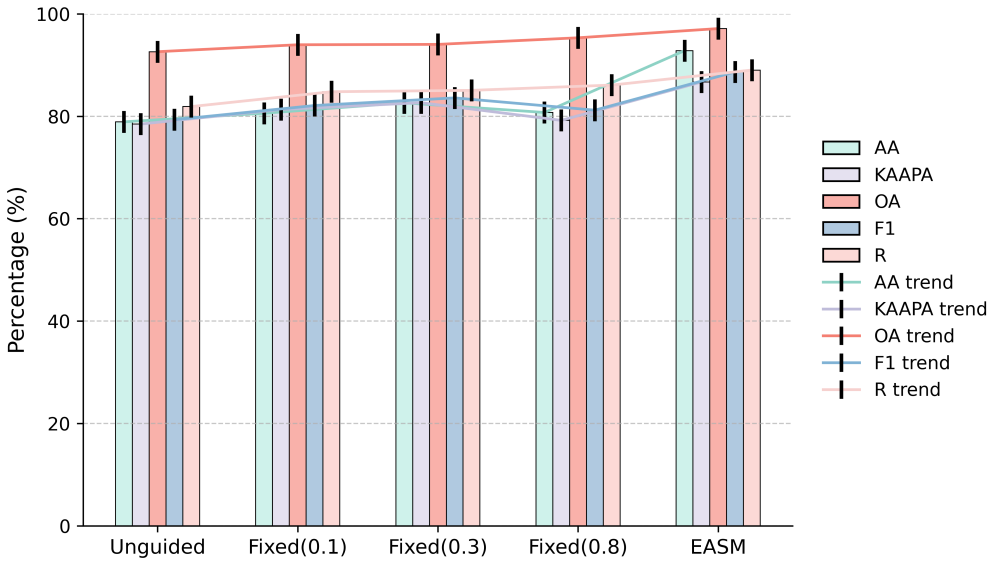


Figure 11. Radar chart comparing segmentation performance under different edge guidance strategies: unguided, fixed edge guidance, and EASM.

5.4. Qualitative Analysis and Limitations

Although GCF-Net achieves state-of-the-art performance on quantitative metrics and demonstrates robust generalization capabilities, we explicitly analyze specific failure cases to provide an objective assessment of its limitations. Figure 12 illustrates representative scenarios corresponding to distinct error patterns.

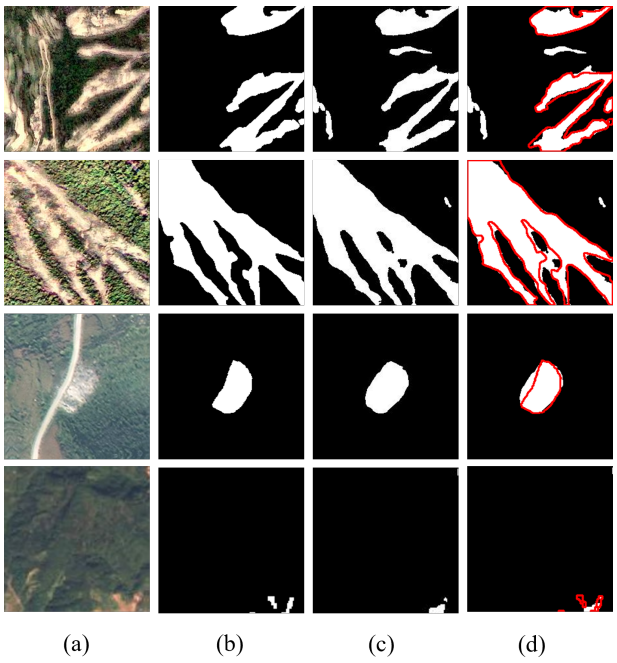


Figure 12. Visualization of typical failure cases. The columns represent: (**a**) original images, (**b**) ground truth, (**c**) prediction results, and (**d**) overlay analysis. In column (**d**), the red outlines denote the ground truth, and the white fills denote the predicted landslides.

Specifically, challenges persist in scenarios characterized by intrinsic ambiguity and complex transition zones. In complex backgrounds, certain non-landslide areas share high visual similarity with landslide textures, making them difficult to distinguish based solely on visual features. This objective similarity can lead to occasional false positives

in non-landslide regions. Similarly, in gradual transition zones, the boundaries between landslides and the background are often blurred rather than sharp. The lack of distinct edge gradients in these low-contrast regions complicates the strict separation of foreground from background, resulting in observable deviations between the predicted boundaries and the ground truth. These observations indicate that accurate segmentation remains challenging in areas with high visual confusion and ambiguous semantic discontinuities.

Regarding boundary irregularity in complex topography, the model sometimes struggles to delineate perfectly sharp edges in regions with intricate topographic changes. As seen in the second row, while the general shape is captured, the predicted boundary may exhibit slight deviations from the ground truth, resulting in minor morphological mismatches along complex geometric margins.The omission of extremely small targets is also observed.The fourth row demonstrates a case of missed detection for a tiny landslide. Despite the multi-scale design of our PGF module, the visual features of such minute targets inevitably diminish during the deep down-sampling operations of the encoder, making them difficult to recover in the decoding stage compared to distinct, large-scale targets.

In summary, while these limitations affect a small fraction of difficult samples, GCF-Net maintains high precision and structural integrity in the vast majority of test cases. Future work will focus on integrating multimodal data to distinguish landslides from confusing backgrounds and exploring refined multi-scale strategies to further enhance the detection capabilities for minute targets.

6. Conclusions

This study proposed GCF-Net, a dual-domain network architecture specifically engineered to break through the accuracy bottleneck of landslide boundary segmentation in complex terrains. To address the limitations of existing methods that process frequency information in a fixed manner or rely solely on single-dimension enhancement, we established a progressive framework that deeply synergizes frequency-domain features with multi-scale spatial semantics. The architecture is founded on three integral components. The Pyramid Lightweight Fusion (PGF) Block combines lightweight convolution with pyramid pooling to construct a compact, high-efficiency hierarchical semantic foundation. The Geometric Context and Frequency Domain Fusion (GCF) module functions as the core module to adaptively fuse dynamic high-frequency masks with spatial context, thereby resolving the dilemma between precise localization and global semantic integrity. Furthermore, the Edge-Aware Detail Consistency Improvement (EDCI) module utilizes boundary confidence guidance to adaptively balance fine detail preservation in edge regions with noise suppression in background areas.

Experimental results on the Hybrid and LMHLD datasets demonstrate that GCF-Net achieves state-of-the-art performance, consistently outperforming representative advanced methods such as SegMAN and RS3Mamba. Specifically, on the Hybrid dataset, our model achieves an OA of 96.42% and AA of 90.28%, surpassing the second-best method, SegMAN, by margins of 1.13% and 1.42%, respectively. Similarly, on the LMHLD dataset, GCF-Net exhibits a remarkable lead in class-balanced accuracy, outperforming SegMAN with a 3.04% increase in AA. Crucially, regarding the boundary quality analysis, GCF-Net demonstrates intrinsic superiority. While recent SSM-based methods like SegMAN and RS3Mamba excel at capturing long-range dependencies, their reliance on patch-based serialization mechanisms may inadvertently limit the reconstruction of fine-grained local details. In contrast, GCF-Net addresses this limitation by explicitly modeling high-frequency components. This advantage is substantiated by quantitative improvements in boundary-specific metrics on the Hybrid dataset, where GCF-Net surpasses SegMAN by 6.32% in P-IoU and 5.65% in B-F1, while reducing the prediction error margin by 5.07 pixels in HD95, confirming

that the proposed framework successfully sharpens the segmentation of complex landslide margins compared to the competitors.

Despite these advancements, the model still faces challenges in specific scenarios. False positives may occur in non-landslide areas that share high visual similarity with landslides, such as bare soil and unpaved roads. Additionally, extremely minute targets are occasionally missed as their visual features inevitably diminish during deep down-sampling operations. Future work will focus on integrating multimodal data to resolve visual ambiguity and exploring refined multi-scale strategies to further enhance the detection capabilities of minute targets.

Author Contributions: Conceptualization, C.D.; investigation, C.D.; writing-original draft preparation, C.D.; data curation, C.D.; methodology, C.D.; formal analysis, Z.L., Y.C., X.Y. and L.Z.; supervision, S.Q. and L.W.; writing-review and editing, S.Q. and L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Heilongjiang Province of China under Grant TD2023D005.

Data Availability Statement: The data and materials supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the anonymous reviewers and editors for their valuable comments and constructive suggestions.

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

References

1. Li, Y.; Fu, B.; Yin, Y.; Hu, X.; Wang, W.; Wang, W.; Li, X.; Long, G. Review on the artificial intelligence-based methods in landslide detection and susceptibility assessment: Current progress and future directions. *Intell. Geoengin.* **2024**, *1*, 1–18. [CrossRef]
2. Scaioni, M.; Longoni, L.; Melillo, V.; Papini, M. Remote sensing for landslide investigations: An overview of recent achievements and perspectives. *Remote Sens.* **2014**, *6*, 9600–9652. [CrossRef]
3. Zhao, C.; Lu, Z.; Zhang, Q.; de La Fuente, J. Large-area landslide detection and monitoring with ALOS/PALSAR imagery data over Northern California and Southern Oregon, USA. *Remote Sens. Environ.* **2012**, *124*, 348–359. [CrossRef]
4. Lin, H.; Huang, H.; Lv, Y.; Du, X.; Yi, W. Micro-UAV based remote sensing method for monitoring landslides in Three Gorges Reservoir, China. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016 ; pp. 4944–4947. [CrossRef]
5. He, H.; Ming, Z.; Zhang, J.; Wang, L.; Yang, R.; Chen, T.; Zhou, F. Robust Estimation of Landslide Displacement From Multitemporal UAV Photogrammetry-Derived Point Clouds. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 6627–6641. [CrossRef]
6. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]
7. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
8. Zhang, Q.; Wang, T. Deep learning for exploring landslides with remote sensing and geo-environmental data: Frameworks, progress, challenges, and opportunities. *Remote Sens.* **2024**, *16*, 1344. [CrossRef]
9. Zhang, H.; Li, L.; Xie, X.; He, Y.; Ren, J.; Xie, G. Entropy guidance hierarchical rich-scale feature network for remote sensing image semantic segmentation of high resolution. *Appl. Intell.* **2025**, *55*, 528. [CrossRef]
10. 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.
11. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation . In *Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022*; Springer: Cham, Switzerland, 2022; pp. 205–218.
12. Ma, X.; Zhang, X.; Pun, M.O. Rs 3 mamba: Visual state space model for remote sensing image semantic segmentation. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6011405. [CrossRef]

13. Fu, Y.; Lou, M.; Yu, Y. SegMAN: Omni-scale context modeling with state space models and local attention for semantic segmentation. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 19077–19087.
14. Casagli, N.; Intrieri, E.; Tofani, V.; Gigli, G.; Raspini, F. Landslide detection, monitoring and prediction with remote-sensing techniques. *Nat. Rev. Earth Environ.* **2023**, *4*, 51–64. [CrossRef]
15. Zhang, Z.; He, S.; Li, Q. Analyzing high-frequency seismic signals generated during a landslide using source discrepancies between two landslides. *Eng. Geol.* **2020**, *272*, 105640. [CrossRef]
16. Strząbała, K.; Ćwiąkała, P.; Puniach, E. Identification of landslide precursors for early warning of hazards with remote sensing. *Remote Sens.* **2024**, *16*, 2781. [CrossRef]
17. Rahman, M.M.; Marculescu, R. G-cascade: Efficient cascaded graph convolutional decoding for 2d medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–7 January 2024; pp. 7728–7737.
18. Azad, R.; Arimond, R.; Aghdam, E.K.; Kazerouni, A.; Merhof, D. Dae-former: Dual attention-guided efficient transformer for medical image segmentation. In *Proceedings of the International Workshop on Predictive Intelligence in Medicine, Vancouver, BC, Canada, 8 October 2023*; Springer: Cham, Switzerland, 2023; pp. 83–95.
19. Yu, C.; Xiao, B.; Gao, C.; Yuan, L.; Zhang, L.; Sang, N.; Wang, J. Lite-hrnet: A lightweight high-resolution network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 10440–10450.
20. Zhou, Y.; Zhang, M.; Wang, Y. Global-Frequency-Domain Network: A Semantic Segmentation Method for High-Resolution Remote Sensing Images Based on Fine-Grained Feature Extraction and Global Context Integration. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 16320–16333. [CrossRef]
21. Yu, J.; Li, X.; Zhang, Z.; Lian, J.; Sun, Y. BS-Mamba: A battery-specific mamba network for robust battery electrode CT image segmentation. *Measurement* **2025**, *259*, 119496. [CrossRef]
22. Qin, Z.; Zhang, P.; Wu, F.; Li, X. Fcanet: Frequency channel attention networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 783–792.
23. Chi, L.; Jiang, B.; Mu, Y. Fast fourier convolution. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 4479–4488.
24. Chen, Y.; Fan, H.; Xu, B.; Yan, Z.; Kalantidis, Y.; Rohrbach, M.; Yan, S.; Feng, J. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3435–3444.
25. Chen, L.; Fu, Y.; Gu, L.; Yan, C.; Harada, T.; Huang, G. Frequency-Aware Feature Fusion for Dense Image Prediction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 10763–10780. [CrossRef]
26. Yang, Y.; Yuan, G.; Li, J. SFFNet: A wavelet-based spatial and frequency domain fusion network for remote sensing segmentation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 3000617. [CrossRef]
27. Yao, T.; Pan, Y.; Li, Y.; Ngo, C.W.; Mei, T. Wave-vit: Unifying wavelet and transformers for visual representation learning. In *Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022*; Springer: Cham, Switzerland, 2022; pp. 328–345.
28. Rao, Y.; Zhao, W.; Zhu, Z.; Zhou, J.; Lu, J. GFNet: Global filter networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 10960–10973. [CrossRef] [PubMed]
29. Clare, M.A.; Lintern, G.; Pope, E.; Baker, M.; Ruffell, S.; Zulkifli, M.Z.; Simmons, S.; Urlaub, M.; Belal, M.; Talling, P.J. Seismic and acoustic monitoring of submarine landslides: Ongoing challenges, recent successes, and future opportunities. In *Noisy Oceans: Monitoring Seismic and Acoustic Signals in the Marine Environment*; Wiley: Hoboken, NJ, USA, 2024; pp. 59–82.
30. Zang, L.; Li, Y. Multi-scale frequency domain learning for texture classification. *Int. J. Mach. Learn. Cybern.* **2025**, *16*, 947–958. [CrossRef]
31. Guo, Y.; Zhao, Y.H. FDMDyG: Frequency Domain-Driven Multi-scale Continuous-Time Dynamic Graph Representation Learning. *Knowl.-Based Syst.* **2025**, *333*, 115045. [CrossRef]
32. Chen, W.; Chen, Z.; Song, D.; He, H.; Li, H.; Zhu, Y. Landslide detection using the unsupervised domain-adaptive image segmentation method. *Land* **2024**, *13*, 928. [CrossRef]
33. NaliniPriya, G.; Lydia, E.L.; Alshenaifi, R.; Kavuri, R.; Ishak, M.K. A two-tiered bidirectional atrous spatial pyramid pooling-based semantic segmentation model for landslide classification using remote sensing images. *IEEE Access* **2024**, *12*, 181316–181331. [CrossRef]
34. Yan, L.; Liu, D.; Xiang, Q.; Luo, Y.; Wang, T.; Wu, D.; Chen, H.; Zhang, Y.; Li, Q. PSP net-based automatic segmentation network model for prostate magnetic resonance imaging. *Comput. Methods Programs Biomed.* **2021**, *207*, 106211. [CrossRef]
35. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
36. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [CrossRef]

37. 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), Munich, Germany, 8–14 September 2018; pp. 3–19.
38. Li, X.; Wang, W.; Hu, X.; Yang, J. Selective kernel networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 510–519.
39. Wang, W.; Xie, E.; Li, X.; Fan, D.P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pvt v2: Improved baselines with pyramid vision transformer. *Comput. Vis. Media* **2022**, *8*, 415–424. [CrossRef]
40. Yu, W.; Luo, M.; Zhou, P.; Si, C.; Zhou, Y.; Wang, X.; Feng, J.; Yan, S. Metaformer is actually what you need for vision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 10819–10829.
41. Guo, M.H.; Lu, C.Z.; Hou, Q.; Liu, Z.; Cheng, M.M.; Hu, S.m. SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation. In *Advances in Neural Information Processing Systems, Proceedings of the 36th International Conference on Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022*; Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2022; Volume 35, pp. 1140–1156.
42. Zhang, W.; Huang, Z.; Luo, G.; Chen, T.; Wang, X.; Liu, W.; Yu, G.; Shen, C. Topformer: Token pyramid transformer for mobile semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 12083–12093.
43. Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; Jiao, J.; Liu, Y. VMamba: Visual State Space Model. In *Advances in Neural Information Processing Systems, Proceedings of the 38th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 10–15 December 2024*; Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2024; Volume 37, pp. 103031–103063. [CrossRef]
44. Guibas, J.; Mardani, M.; Li, Z.; Tao, A.; Anandkumar, A.; Catanzaro, B. Adaptive fourier neural operators: Efficient token mixers for transformers. *arXiv* **2021**, arXiv:2111.13587.
45. Han, D.; Pan, X.; Han, Y.; Song, S.; Huang, G. Flatten transformer: Vision transformer using focused linear attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 5961–5971.
46. Tang, Y.; Han, K.; Guo, J.; Xu, C.; Li, Y.; Xu, C.; Wang, Y. An image patch is a wave: Phase-aware vision mlp. In Proceedings of the IEEE/CVF Conference on Computer vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 10935–10944.
47. Li, X.; You, A.; Zhu, Z.; Zhao, H.; Yang, M.; Yang, K.; Tan, S.; Tong, Y. Semantic flow for fast and accurate scene parsing. In *Proceedings of the European Conference on Computer Vision, Virtual, 23–28 August 2020*; Springer: Cham, Switzerland, 2020; pp. 775–793.
48. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 5728–5739.
49. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
50. Si, C.; Huang, Z.; Jiang, Y.; Liu, Z. Freeu: Free lunch in diffusion u-net. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 4733–4743.
51. Kirillov, A.; Wu, Y.; He, K.; Girshick, R. Pointrend: Image segmentation as rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 9799–9808.
52. Yuan, Y.; Xie, J.; Chen, X.; Wang, J. Segfix: Model-agnostic boundary refinement for segmentation. In *Proceedings of the European Conference on Computer Vision, Virtual, 23–28 August 2020*; Springer: Cham, Switzerland, 2020; pp. 489–506.
53. Borse, S.; Wang, Y.; Zhang, Y.; Porikli, F. Inverseform: A loss function for structured boundary-aware segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 5901–5911.
54. Lazarow, J.; Xu, W.; Tu, Z. Instance segmentation with mask-supervised polygonal boundary transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 4382–4391.
55. Xu, J.; Xiong, Z.; Bhattacharyya, S.P. PIDNet: A real-time semantic segmentation network inspired by PID controllers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 19529–19539.
56. Zhang, G.; Lu, X.; Tan, J.; Li, J.; Zhang, Z.; Li, Q.; Hu, X. Refinemask: Towards high-quality instance segmentation with fine-grained features. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 6861–6869.
57. Ke, L.; Danelljan, M.; Li, X.; Tai, Y.W.; Tang, C.K.; Yu, F. Mask transfiner for high-quality instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–24 June 2022; pp. 4412–4421.
58. Huang, S.; Lu, Z.; Cheng, R.; He, C. FaPN: Feature-aligned pyramid network for dense image prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 864–873.

59. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. Carafe: Content-aware reassembly of features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3007–3016.
60. Liu, W.; Lu, H.; Fu, H.; Cao, Z. Learning to upsample by learning to sample. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Vancouver, BC, Canada, 18–22 June 2023; pp. 6027–6037.
61. Ma, A.; Wang, J.; Zhong, Y.; Zheng, Z. FactSeg: Foreground activation-driven small object semantic segmentation in large-scale remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5606216. [CrossRef]
62. Tsai, F.J.; Peng, Y.T.; Tsai, C.C.; Lin, Y.Y.; Lin, C.W. BANet: A blur-aware attention network for dynamic scene deblurring. *IEEE Trans. Image Process.* **2022**, *31*, 6789–6799. [CrossRef]
63. Wang, L.; Li, R.; Zhang, C.; Fang, S.; Duan, C.; Meng, X.; Atkinson, P.M. UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 196–214. [CrossRef]
64. 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 14–19 June 2020; pp. 11534–11542.
65. 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.
66. Zhao, R.; Qian, B.; Zhang, X.; Li, Y.; Wei, R.; Liu, Y.; Pan, Y. Rethinking dice loss for medical image segmentation. In *Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 17–20 November 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 851–860.
67. Ji, S.; Yu, D.; Shen, C.; Li, W.; Xu, Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **2020**, *17*, 1337–1352. [CrossRef]
68. Ghorbanzadeh, O.; Xu, Y.; Zhao, H.; Wang, J.; Zhong, Y.; Zhao, D.; Zang, Q.; Wang, S.; Zhang, F.; Shi, Y.; et al. The outcome of the 2022 Landslide4Sense competition: Advanced landslide detection from multisource satellite imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9927–9942. [CrossRef]
69. Liu, G.; Wang, Y.; Chen, X.; Du, B.; Li, P.; Wu, Y.; Fang, Z.; Ma, P. LMHLD: A Large-Scale Multisource High-Resolution Landslide Dataset for Landslide Detection Based on Deep Learning. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5648415. [CrossRef]
70. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv* **2021**, arXiv:2102.04306. [CrossRef]
71. Rahman, M.M.; Marculescu, R. Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation. In Proceedings of the 7nd International Conference on Medical Imaging with Deep Learning, Paris, France, 3–5 July 2024; pp. 1526–1544.

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.