SCGC-Net:Spatial Context Guide Calibration Network for multi-source RSI Landslides detection

Yukun Fan, Peifeng Ma, *Senior Member, IEEE*, Qingbo Hu, Guiwei Liu, Zihuan Guo, Yixian Tang, Fan Wu, *Member, IEEE*, and Hong Zhang, *Member, IEEE*

***Abstract*—Landslide is a common geological disaster, and rapid landslide extraction using high-resolution remote sensing imagery (RSI) is of great significance for emergency rescue and damage assessment. In RSI, landslides often have irregular shapes, large scale variations, and are easily affected by environmental factors. Existing deep learning methods have limited ability in extracting multi-scale features, integrating these features effectively, and adapting to complex environments, resulting in models that are not optimized for robustness. To overcome these challenges, this study proposes a spatial context guided calibration network (SCGC-Net) for multi-source remote sensing data. SCGC-Net introduces a novel combination of hybrid multi-scale feature extraction, context-aware modulation of landslide characteristics, and a progressive feature calibration fusion strategy, enabling efficient feature extraction, accurate feature integration, and enhanced cross-domain generalization when working with multi-source remote sensing data. SCGC-Net was tested on several datasets representing diverse geographical regions and imaging platforms, including the CAS Landslide Dataset, HR-GLDD, Bijie and GVLM. Experimental results indicate that SCGC-Net outperforms existing methods across all evaluation metrics and exhibits superior generalization performance in domain adaptation experiments.**

***Index Terms*—Landslide detection, deep learning, High-resolution remote sensing imagery.**

I. Introduction

LANDSLIDES represent a frequent and highly destructive natural hazard, posing significant risks to human life, property, and infrastructure. With the intensification of global climate change and the accelerated process of urbanization, both the frequency and severity of landslide occurrences are increasing [1]. Consequently, the timely and precise detection of landslide-affected areas is critical for emergency response, disaster assessment, and post-disaster recovery.

Traditional methods of landslide detection primarily rely on field surveys and visual interpretation; however, these approaches are often time-intensive and laborious, limiting their effectiveness in providing rapid large-scale response. As Earth observation technologies have advanced, landslide detection using high-resolution remote sensing imagery has emerged as a prominent area of research [2]–[4]. Optical remote sensing data, particularly high-resolution imagery from satellite and airborne platforms, offer an ideal source for detecting landslides. These data not only provide high spatial and temporal resolution but also rich spectral information, which aids in identifying the subtle features of landslide events [5].

In recent years, deep learning techniques have demonstrated significant advantages in the rapid detection of landslide disasters due to their high degree of automation, advanced feature extraction capabilities, and end-to-end learning processes [6]. These models have shown superior performance, particularly when applied to large-scale, high-resolution remote sensing imagery, outperforming traditional methods [7]. Multiple research groups have recently developed high-quality landslide detection datasets, including high-resolution satellite-based datasets [4], [8], [9] and high-resolution aerial imagery datasets [10]. Xu et al. [11] integrated satellite and Unmanned Aerial Vehicle (UAV) data to create a large-scale multi-sensor landslide dataset. These datasets encompass a wide range of geographic regions, landslide triggering mechanisms, and imaging platforms, advancing the development of landslide detection algorithms.

Despite the growing availability of data resources and improved algorithms, landslide detection still faces significant challenges. One key issue is the robustness of models when processing images from diverse remote sensing platforms [11]. This issue becomes particularly pronounced when conducting large-scale assessments that require the integration of multi-source remote sensing data. Therefore, the development of a robust landslide detection method capable of effectively processing multi-source remote sensing data and exhibiting strong generalization capabilities is a pressing need in this field of research.

Landslides exhibit significant scale variance and irregu-

This work was supported in part by the International Research Center of Big Data for Sustainable Development Goals under Grant CBASYX0906, and in part by the Major Science and Technology Projects of China State Railway Group Co. Ltd. under Grants K2023G032 and N2023G072 (Corresponding author: Hong Zhang).

Yukun Fan and Zihuan Guo are with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and with the International Research Center of Big Data for Sustainable Development Goals, Beijing 100049, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: fanyukun22@mails.ucas.ac.cn; guozihuan21@mails.ucas.ac.cn).

PeiFeng Ma is with the Department of Geography and Resource Management and the Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong(e-mail: mapeifeng@cuhk.edu.hk).

Qingbo Hu is with China Railway Economic and Planning Research Institute Co. Ltd., Beijing 100844, China (e-mail: huqingbo@vip.sina.com).

Guiwei Liu is with China Railway Design Corporation, Tianjin 300251, China (e-mail: liuyun5216@163.com).

Yixian Tang, Fan Wu and Hong Zhang are with the International Research Center of Big Data for Sustainable Development Goals, Beijing 100049, China, and with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: tangyx@aircas.ac.cn; wufan@aircas.ac.cn; zhanghong@radi.ac.cn).

lar morphological characteristics in high-resolution remote sensing images, particularly in multi-platform datasets. These characteristics make it essential to simultaneously capture both fine-grained local features and global contextual relationships for accurate segmentation. At present, Convolutional Neural Network (CNN) based semantic segmentation models, such as Fully Convolutional Network (FCN) [12] and U-Net [13], are widely applied to landslide detection tasks [7], [14]–[16]. However, the limited receptive field of CNNs constrains their ability to capture long-range dependencies. To address this limitation, researchers have introduced the self-attention mechanism from Transformer models [17], [18], which enhances global feature extraction but comes with high computational overhead and reduced effectiveness in capturing local details. The recently proposed Mamba state space model (SSM) [19] offers a novel approach to overcoming these limitations. Mamba effectively extracts global features while maintaining linear time complexity and has demonstrated advantages across several computer vision tasks [20]–[22]. However, its potential for landslide segmentation in remote sensing images remains underexplored. A key research direction lies in integrating Mamba's global modeling capabilities with improved local feature extraction to develop a model architecture capable of addressing the multi-scale nature of landslides.

High-resolution remote sensing-based landslide detection faces not only the challenge of multi-scale feature extraction but also the inherent complexity of landslides. Landslides appear with diverse geographical backgrounds and occlusion factors in remote sensing imagery, resulting in complex or blurred boundaries. Various multi-scale feature fusion strategies have been proposed, including skip connections [23], [24], attention mechanisms [25], [26], and image pyramid strategies [27], [28]. Despite their progress in managing multi-scale features and complex backgrounds, these methods have limitations:

(1) Skip connections can introduce shallow-layer noise, leading to semantic inconsistencies between shallow and deep features, which compromise information integration.

(2) Current attention mechanisms and pyramid strategies do not fully resolve spatial misalignment between feature maps, which may result in boundary detail loss and difficulty in aggregating small-scale landslide features.

(3) Existing methods struggle to adapt to variations in geographic environments and imaging conditions, restricting their domain generalization capabilities.

An effective fusion strategy should be capable of addressing multi-scale features, correcting spatial misalignment, and adapting to diverse geographic environments and imaging conditions to improve the accuracy and robustness of landslide detection in complex scenarios.

To address these key challenges, this study introduces the spatial context guided calibration Network (SCGC-Net). SCGC-Net is designed to improve multi-scale feature extraction, feature fusion, and generalization across multi-source data. Its core architecture consists of three innovative modules:

(1) Hybrid Multi-scale Information Extraction (HMIE): This module combines Mamba and CNN structures to form a hybrid feature extractor. By integrating Mamba with dilated convolution, it achieves efficient long-range dependency modeling through Mamba, while the dilated convolution branch enhances local spatial feature extraction. This design improves the model's ability to capture multi-scale landslide features while maintaining efficient time complexity.

(2) Context-Aware Modulator (CAM): A multi-level context-aware mechanism, coupled with a gating strategy, dynamically fuses multi-scale contextual information across spatial and channel dimensions, enhancing the model's generalization performance across various platforms and geographic regions.

(3) Progressive Spatial-Context Calibration Strategy (PSCCS): By learning pixel-level calibration offsets, this module achieves precise calibration and fusion of feature maps at different resolutions, effectively mitigating information loss during downsampling and improving the accuracy of boundary segmentation and small-scale landslide detection.

This study comprehensively evaluated SCGC-Net on three representative datasets: the CAS Landslide Dataset (CLD) [11], HR-GLDD [8], and Bijie Dataset [4]. SCGC-Net achieved the best performance across nearly all accuracy metrics. On the CLD, SCGC-Net achieved an intersection over union (IoU) of 87.85%, surpassing the second-best method by 2.30%. In the HR-GLDD dataset, SCGC-Net exceeded the second-best method by 1.88% in Recall and 2.61% in IoU. For the Bijie dataset, SCGC-Net achieved an F1 score of 84.75% and an IoU of 73.53%, representing improvements of 2.08% and 3.06% over the best comparison method. Additionally, in the generalization experiments conducted on the global very-high-resolution landslide mapping (GVLM) dataset [9], SCGC-Net exhibited exceptional domain adaptability, maintaining top performance across different geographical environments and imaging conditions. These results robustly demonstrate SCGC-Net's effectiveness and generalization ability in handling multi-source, multi-scale landslide detection tasks, offering valuable technical support for improving disaster response and developing a universal landslide detection system.

II. Related Work

A. Feature Extraction Techniques For Landslide Segmentation

Early studies on feature extraction techniques for landslide segmentation predominantly utilized CNN-based models such as FCN [12] and U-Net [13]. Liu et al. [29] enhanced the extraction of multi-scale landslide features by introducing channel attention mechanisms and Atrous Spatial Pyramid Pooling (ASPP). Li et al. [30] combined Faster Region-based Convolutional Neural Network (Faster-RCNN) [31] with edge detection algorithms to improve U-Net's capability in extracting landslide edge features. HADeenNet [25] boosted the performance of small-scale landslide segmentation by parallel processing of input images at different resolutions. However, the local receptive field inherent in CNNs limits their capacity to capture long-range dependencies, which are critical for modeling global contextual relationships in landslide detection. To address this issue, researchers have started exploring alternative methods. Lu et al. [32] proposed a multi-task learning approach that integrates object classification with semantic segmentation networks to simultaneously optimize global and local feature extraction. To

overcome CNN's limitations, Transformer-based architectures have been introduced. Fu et al. [33] and Ghorbanzadeh et al. [34] implemented Swin Transformer [35] and SegFormer [36] models, which utilize hierarchical self-attention mechanisms to enhance the extraction of complex features in landslide segmentation. Additionally, Lv et al. [17] and Huang et al. [18] integrated Transformer models with morphological edge extraction techniques, thereby improving the model's ability to perceive landslide shapes and boundaries.

Given the complementary strengths of CNNs and Transformers in capturing local and global features, respectively, researchers have begun developing hybrid structures. For instance, Li et al. [37] fused CNN-based feature extraction branches with Transformer-based global feature extraction branches, while Wu et al. [38] combined shallow CNN layers with deeper Swin Transformer layers, both of which improved the extraction of multi-scale landslide features. However, these hybrid approaches still face key challenges: Transformer models suffer from quadratic growth in computational complexity with increasing sequence length, which leads to inefficiencies when processing high-resolution remote sensing images. While CNNs excel at capturing local features, they are less effective at modeling global context. Recently, Gu et al. [19] proposed the Mamba model, which addresses the efficiency concerns of Transformers by capturing long-range dependencies with linear complexity. Nonetheless, the capability of Mamba to maintain spatial structure and extract local features remains underexplored. This inspires further research into combining Mamba's global modeling efficiency with the spatial detail extraction capabilities of CNNs to develop a solution that balances computational efficiency with comprehensive feature extraction for landslide segmentation.

B. Feature Fusion Techniques For Landslide Segmentation

Feature fusion techniques are crucial in landslide segmentation tasks, as effective fusion strategies help to better integrate semantic information, thus improving segmentation accuracy. Many researchers have applied U-Net's [13] skip connection mechanism to landslide detection, integrating deep and shallow features to mitigate the information loss caused by FCN's direct downsampling and upsampling operations [30], [39], [40]. However, simple feature concatenation can lead to inconsistencies between feature scales, which can negatively impact the overall fusion effectiveness. To better accommodate the multi-scale variability of landslide features, researchers have developed various multi-scale feature fusion strategies. AMU-Net [41] introduced multi-scale modules within skip connections, incorporating additional contextual information when fusing deep and shallow features. Zheng et al. [27] employed the DeepLabV3+ [42] model, which fuses multi-scale contextual information through ASPP, using parallel dilated convolutions with varying dilation rates. GMNet [28] combines the concepts of Feature Pyramid Networks (FPN) [43] and Pyramid Pooling Modules (PPM) [44], proposing a multi-scale feature fusion module. By performing top-down and bottom-up feature fusion, the model enhances its ability to detect multi-scale landslides. Wu et al. [38] proposed a multi-branch feature fusion approach that integrates encoder features using depthwise separable convolution, dilated convolution, and 1×1 convolution, followed by element-wise addition for fusion. This method aims to simultaneously capture spatial, spectral, and multi-scale features. While these multi-scale fusion techniques have improved model performance, they often overlook the interrelationships between features at different scales, potentially leading to feature redundancy.

To address the problem of information redundancy in multi-scale feature fusion and to strengthen the representation of key features, researchers have introduced attention mechanisms into the feature fusion process. HADeenNet [25] adopted a multi-scale feature fusion strategy combined with attention mechanisms, enabling adaptive weighting of features at different scales. DPANet [45] combined pyramid pooling feature fusion with dual attention mechanisms in both spatial and channel dimensions, enhancing multi-scale feature representation while capturing global context and local details. MFFSP [37] incorporated self-attention for global information and convolution for local information into the multi-branch feature fusion process, enhancing the global context understanding in landslide images. However, the high computational complexity of self-attention mechanisms may limit their applicability in high-resolution remote sensing images.Despite the progress made by these feature fusion techniques, they continue to face the challenge of spatial feature misalignment. Simply fusing features at different scales without considering potential spatial inconsistencies can lead to the accumulation of such inconsistencies throughout the network, resulting in blurred boundary features and reduced ability to effectively detect small-scale landslides, significantly impacting detection accuracy. To address this issue, this study proposes PSCCS. By introducing calibration unit (CU) between features of different scales, PSCCS enables adaptive feature adjustment and fusion, effectively alleviating the problem of feature misalignment.

III. Proposed Method

This section presents the overall architecture and workflow of the proposed SCGC-Net. The overall structure of SCGC-Net is illustrated in Fig. 1. During the feature extraction phase, the model utilizes the HMIE structure, developed in this study, which integrates the strengths of both CNN and Mamba to effectively extract information from landslides of varying scales and forms. The CAM module, employing a modulation mechanism, is designed to adaptively refine multi-scale semantic information, serving as a replacement for self-attention to enhance both the accuracy and generalization of semantic recognition. In the final multi-scale feature fusion stage, the PSCCS module precisely calibrates and fuses feature maps at different resolutions, leading to the final output.

A. Mamba Preliminaries

Mamba is an advanced sequence modeling technique based on SSM, designed to effectively capture long-range dependencies and adapt to complex spatio-temporal dynamics. SSM originates from control theory and maps an input sequence $\mathbf{x}(t) \in \mathbb{R}^L$ to an output sequence $\mathbf{y}(t) \in \mathbb{R}^L$ through a latent state $\mathbf{h}(t) \in \mathbb{R}^N$. The fundamental form of the

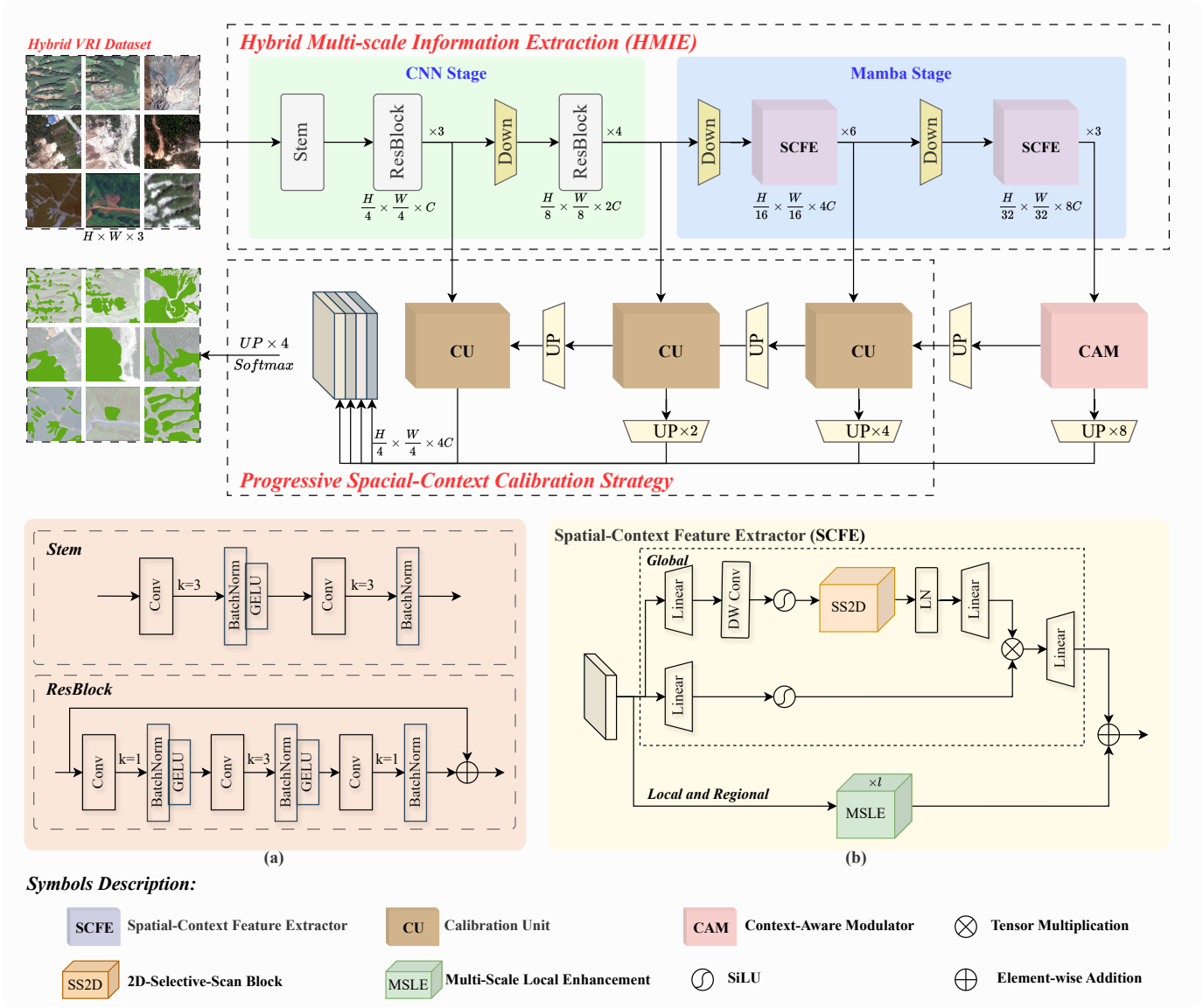

Fig. 1. Architecture illustration of SCGC-Net. The network integrates HMIE, CAM, and PSCCS for effective landslide detection. (a) Detailed structure of CNN stage in feature extraction, showing the Stem and ResBlock modules. (b) Architecture of SCFE in Mamba stage for global-local feature learning.

SSM is governed by the following continuous-time ordinary differential equation:

$$\begin{aligned} \mathbf{h}'(t) &= \mathbf{A}\mathbf{h}(t) + \mathbf{B}\mathbf{x}(t), \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{h}(t). \end{aligned} \tag{1}$$

Here, $\mathbf{A} \in \mathbb{R}^{N\times N}$ is the state transition matrix, which controls the temporal evolution of the hidden state; $\mathbf{B} \in \mathbb{R}^{N\times 1}$ and $\mathbf{C} \in \mathbb{R}^{1\times N}$ represent the input and output projection matrices, respectively. To integrate SSM into deep learning, it is necessary to discretize the continuous-time model. Mamba achieves this using the zero-order hold (ZOH) method, introducing a time scale parameter Δ. The discretized SSM is expressed as:

$$\begin{aligned} \mathbf{h_t} &= \bar{\mathbf{A}}\mathbf{h_{t-1}} + \bar{\mathbf{B}}\mathbf{x_t}, \\ \mathbf{y_t} &= \bar{\mathbf{C}}\mathbf{h_t}. \end{aligned} \tag{2}$$

where $\bar{\mathbf{A}} = e^{\Delta\mathbf{A}}$, $\bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(e^{\Delta\mathbf{A}} - \mathbf{I})\Delta\mathbf{B}$, and $\bar{\mathbf{C}} = \mathbf{C}$.

Mamba's unique feature is its selective scanning mechanism (S6), which dynamically adjusts the parameters $\mathbf{B}$, $\mathbf{C}$, and Δ based on the input data. This allows the model to adapt to varying input characteristics and contexts. Specifically, for an input $\mathbf{x} \in \mathbb{R}^{B\times L\times D}$, the parameters $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{B\times L\times N}$ and $\Delta \in \mathbb{R}^{B\times L\times D}$ are dynamically modified, where B is the batch size, L is the sequence length, D is the feature dimension, and N is the hidden state size.

B. Hybrid Multi-scale Information Extraction

We propose a novel feature extraction method, HMIE, which effectively combines residual convolutional networks with an improved Mamba structure for comprehensive and efficient landslide image feature extraction. This design captures both local details and long-range dependencies, significantly enhancing the accuracy and robustness of landslide detection.

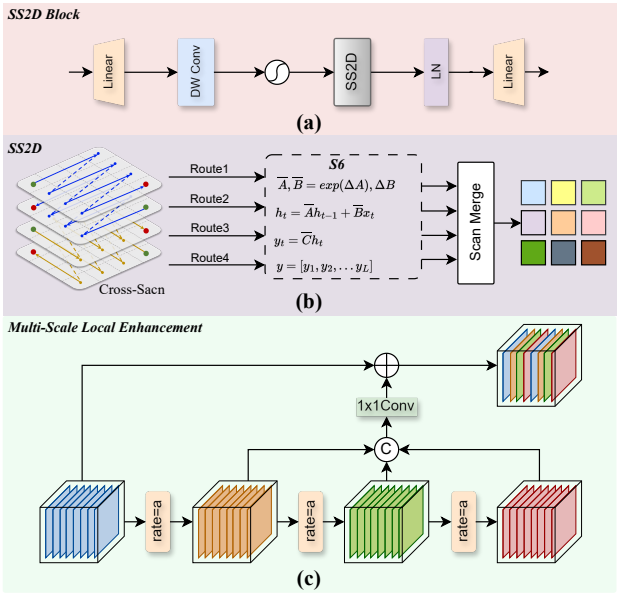

Fig. 2. Key components of SCFE module in SCGC-Net. (a) Architecture of SS2D Block. (b) Illustration of 2D-Selective-Scan (SS2D). (c) Architecture of MSLE module.

HMIE is designed to take advantage of the complementary strengths of CNN and Mamba at different levels of the network. Specifically, HMIE includes the following key components: 1. CNN residual blocks: Used in the early stages to extract local features.
2. Spatial context dual-branch module: Consists of a context branch based on the improved Mamba structure and a spatial branch using dilated convolution.
This design enables efficient extraction of local features in the shallow layers while capturing global contextual information in the deeper layers, allowing for multi-scale, comprehensive extraction of landslide characteristics.

For an input image of size $H \times W \times C$, the image is first downsampled using a stem block composed of CNN convolutional layers, generating a 2D feature map with a resolution of $\frac{H}{4} \times \frac{W}{4}$. Next, we construct the CNN residual blocks using an improved ResNet50 [46] bottleneck architecture with the GELU [47] activation function, as outlined below:

$$z_1 = \text{GELU}(\text{BN}(\text{Conv}_{1\times1}(x))). \tag{3}$$

$$z_2 = \text{GELU}(\text{BN}(\text{Conv}_{3\times3}(z_1))). \tag{4}$$

$$z_3 = \text{BN}(\text{Conv}_{1\times1}(z_2)). \tag{5}$$

$$\text{output} = \text{GELU}(z_3 + x). \tag{6}$$

Here, GELU refers to the Gaussian Error Linear Unit activation function, and BN stands for batch normalization. This architecture effectively extracts local spatial features, establishing a strong foundation for subsequent global feature extraction stages.

To comprehensively capture the multi-scale features of landslide images, we designed a Spatial-Context Feature Extractor (SCFE), as shown in Fig. 1 (b). This module consists of two branches: a context branch to capture global context information and a spatial branch to capture local spatial features.

Context Branch: Inspired by the vanilla Visual State Space (VSS) architecture [20] and Vision Mamba (Vim) [48], the context branch features a dual-path structure to efficiently capture long-range dependencies in landslide images. Both paths start from the same input feature x, but each undergoes different processing. The first path is simple, consisting of a linear transformation layer followed by an activation function, and primarily retains and slightly transforms the original feature information:

$$\text{Path}_1(x) = \sigma(\text{Linear}(x)). \tag{7}$$

The second path, which uses the Mamba structure, is designed to model more complex long-range contextual information:

$$\text{Path}_2(x) = \text{Linear}(\text{LN}(\text{SS2D}(\sigma(\text{DWConv}(\text{Linear}(x)))))). \tag{8}$$

Here, σ denotes the SiLU [49] activation function, LN refers to layer normalization, and DWConv represents depthwise separable convolution. The outputs of the two paths are combined using elementwise multiplication to produce the global contextual feature F_{global}. This architecture facilitates the preservation of original feature information while enabling efficient extraction of higher-level feature representations, thus enhancing the model's perception and understanding of landslide features.

The core of the context branch is the 2D-Selective-Scan (SS2D) module, whose processing flow is illustrated in Fig. 2 (a). The SS2D process is defined as follows:

$$\text{SS2D}(\mathbf{X}) = \mathbf{W}_2 \cdot \text{LN}(\text{Scan}(\sigma(\text{DWConv}(\mathbf{W}_1 \cdot \mathbf{X})))). \tag{9}$$

Here, $\mathbf{X} \in \mathbb{R}^{B\times L\times D}$ represents the input feature, where B is the batch size, L is the sequence length, and D is the feature dimension. The matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are linear transformation matrices for the input and output, respectively.

The Scan operation extends the traditional bidirectional scan in vision mamba to a four-directional cross-scan strategy, as shown in Fig. 2 (b). This includes both horizontal and vertical bidirectional scans. Each scan path performs state-space modeling, and the results of the four scan paths are merged using S6 and scan merge operations to produce the final output features:

$$\text{Scan}(\mathbf{X}) = \sum_{d=1}^{4} \mathbf{Y}^d. \tag{10}$$

This design enables the context Mamba branch to capture long-range dependencies from multiple directions, effectively integrating information from various spatial locations and improving the model's ability to perceive global context.

Spatial Branch: While the SSM module provides global receptive field modeling, its serial processing may result in the loss of local detail features, which can lead to insufficient representation of regions and local details. This is especially important for the precise modeling of landslide features. To

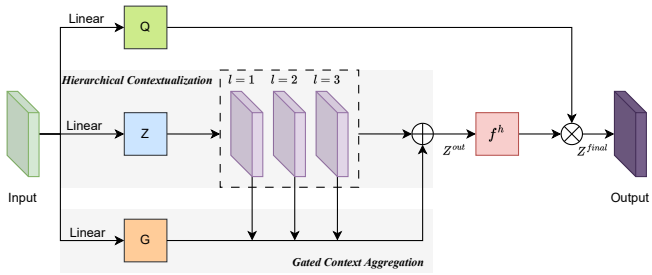

Fig. 3. Context-Aware Modulator

address this, the spatial branch introduces a Multi-Scale Local Enhancement (MSLE) module to compensate for the potential omission of local detail features by the context branch. The MSLE consists of three dilated convolution layers with different dilation rates, as illustrated in Fig. 2 (c). The feature maps from these layers, each covering a distinct local receptive field, are concatenated. A 1×1 convolution is then applied to reduce the dimensionality. Finally, the downsampled features are connected with the input features via a residual connection, allowing the output feature map to integrate spatial features from three different local receptive fields. This enhances the model's ability to capture fine-grained local features. The dilated convolution is expressed as follows:

$$Z(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} W(m,n) \cdot X(i + m \cdot d, j + n \cdot d). \quad (11)$$

where d is the dilation rate, and W is the convolution kernel. The MSLE operation is defined as:

$$\mathrm{MSLE(X)} = \mathrm{C(D_a(X), D_b(X), D_c(X))} + \mathrm{X}. \quad (12)$$

where D_a, D_b, and D_c represent dilated convolution operations with different dilation rates, and C represents the feature concatenation and compression operation.

In conclusion, HMIE successfully integrates CNN and Mamba structures, combining both local feature extraction and global context modeling. The context branch, with its improved SS2D module, captures long-range dependencies through a four-directional cross-scanning approach. Meanwhile, the MSLE module enhances local detail extraction via multi-scale dilated convolutions, compensating for critical local information. This integrated design enables HMIE to efficiently and comprehensively capture the multi-scale features and complex spatial relationships in landslide images, significantly improving detection performance across a variety of challenging scenarios.

C. Context-Aware Modulator

Although the HMIE module effectively extracts multi-scale information of landslides, there are still issues of insufficient semantic refinement and semantic redundancy during feature recovery. To address these problems, we propose the CAM Block, which aims to refine contextual semantic information at different scales and aggregate the most instructive context for classification. Fig. 3 shows the overall architecture of CAM.

Traditional methods typically treat contextual information from all scales as equally important and aggregate context within a predefined range. Since large-scale landslides contain more landslide pixels, small-scale landslides are smoothed into other land features during the semantic information aggregation process, resulting in a semantic bias toward large-scale landslide areas and causing context mismatches.To overcome these limitations, CAM introduces the concept of focal modulation based on the work by Yang et al. [50], which enables adaptive focusing and enhancement of contextual information at different scales. First, the feature $\mathbf{X}$ from the backbone network is projected via a linear transformation to enhance the discriminative power of the input features:

$$\mathbf{Z}^0 = f^Z(\mathbf{X}) \in \mathbb{R}^{H \times W \times C}. \quad (13)$$

where f^Z is the linear projection function, and H, W, and C represent the height, width, and channel dimensions of the feature map, respectively. The projected feature $\mathbf{Z}^0$ is then passed through N depthwise convolution layers to extract contextual information at different granularity levels:

$$\mathbf{Z}^n = (f_a^n(\mathbf{Z}^{n-1}) = \mathrm{SiLU(DWConv}(\mathbf{Z}^{n-1}))) \in \mathbb{R}^{H \times W \times C}. \quad (14)$$

Here, f_a^n denotes the contextual function at layer n. We use the SiLU activation function because it outperformed ReLU [51] and other activation functions in our experiments, particularly when handling complex nonlinear relationships. The convolution kernel size k is initialized to 5 in the first layer and increases by 2 in each subsequent layer. This design ensures that $k^n < k^{n+1}$, forming a hierarchical receptive field structure. The final receptive field size is:

$$r = 1 + \sum_{i=1}^{n} (k^i - 1). \quad (15)$$

This adaptive receptive field design enables the module to capture multi-scale contextual information, ranging from local to global, which is particularly advantageous for detecting landslide regions of varying sizes. To focus on the contextual information most important for pixel-wise semantic classification, we introduce a gated aggregation mechanism, allowing selective, context-aware features to enter subsequent layers. Specifically, given N contextual feature maps from the previous step $\mathbf{Z} = \mathbf{Z}^n$, the final output $\mathbf{Z}_out$ is obtained by performing element-wise multiplication, which results in a weighted sum, similar to the self-attention mechanism:

$$\mathbf{Z}^{out} = \sum_{i=1}^{N} \mathbf{G}^i \odot \mathbf{Z}^i \in \mathbb{R}^{H \times W \times C}. \quad (16)$$

where $\mathbf{G}^i \in \mathbb{R}^{H \times W \times 1}$ represents the gating weights, and $\odot$ denotes element-wise multiplication. This mechanism allows the model to dynamically adjust the importance of contextual representations at different scales, particularly aiding in distinguishing large and small landslides and addressing edge regions. To further enhance feature representation, channel-wise feature fusion is performed after spatial aggregation:

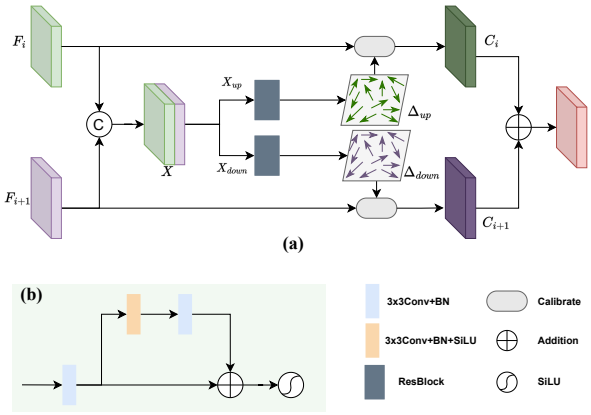

Fig. 4. (a) Illustration of CU. (b) Structure of Residual Block (RB) in CU.

$$\mathbf{Z}^{\text{final}} = f^h(\mathbf{Z}^{\text{out}}). \tag{17}$$

where f^h is the channel fusion function, implemented as a linear transformation. Through this design, CAM can adaptively focus and enhance contextual information at different scales, effectively handling multi-scale landslide features and improving the model's generalization ability across different types of remote sensing images.

D. Progressive Calibration Strategy

In multi-scale feature fusion for landslide detection, spatial inconsistencies and semantic disparities between features of different resolutions pose a common challenge. In structures like skip connections or feature pyramids, there are significant differences in both spatial and semantic properties across various feature levels. Directly fusing these features may lead to information loss and degraded segmentation performance of landslide characteristics [52]. To address this issue while maximizing the semantic value of multi-scale features, we propose the PSCCS.

PSCCS aims to optimize feature fusion across multiple scales by progressively calibrating adjacent features, ensuring effective long-range feature alignment, and simultaneously integrating multi-scale information to enhance feature representations. Unlike traditional methods such as FPN and PPM, PSCCS employs a bidirectional calibration mechanism. This mechanism not only leverages high-level semantic information to guide lower-level features but also refines high-level features using the detailed information from lower-level features. This bidirectional interaction is reasonably effective for addressing the complex terrain and texture details often encountered in landslide detection. The core of PSCCS consists of a series of Calibration Unit (CU), as shown in Fig. 4, each responsible for calibrating features across two adjacent scales. Each CU takes adjacent multi-scale features from HMIE as inputs, denoted as $F_i, F_{i+1} \in \mathbb{R}^{H\times W\times C}$, where F_i represents the upsampled low-resolution feature, and F_{i+1} is the feature from the downsampling stage. The calibration process leverages residual learning to enhance semantic flow estimation.

The CU first concatenates F_i and F_{i+1} along the channel dimension to capture cross-scale feature relationships. Next, the concatenated features are processed through a dual-branch architecture to learn calibration offsets. One branch learns the offsets from high to low resolution (downsampling branch), while the other learns from low to high resolution (upsampling branch). To generate more accurate semantic flow for feature calibration, we use Residual Block (RB) in a dual-branch architecture, which consists of two 3×3 convolutional layers with batch normalization and SiLU activation, connected by a skip connection to maintain feature propagation (see Fig. 4(b)). Compared to direct convolution operations, the residual structure helps better capture semantic correspondences between features,

$$\begin{aligned}\Delta_{\text{down}} &= \text{RB}_{\text{down}}([F_i, F_{i+1}]) \\ \Delta_{\text{up}} &= \text{RB}_{\text{up}}([F_i, F_{i+1}])\end{aligned} \tag{18}$$

where $[\cdot,\cdot]$ denotes channel-wise concatenation, and $\Delta_{\text{down}}, \Delta_{\text{up}}$ represent the learned calibration offsets. Based on these learned offsets, the CU performs bidirectional feature calibration:

$$\begin{aligned}\hat{C}_i &= f(F_{i+1}, \Delta_{\text{up}}) + F_i \\ \hat{C}_{i+1} &= f(F_{i+1}, \Delta_{\text{down}}) + F_{i+1}\end{aligned} \tag{19}$$

where $f(\cdot)$ implements feature calibration through bilinear sampling, and the residual connections preserve original feature characteristics. In the end, the calibrated features $\hat{C}_i$ and $\hat{C}_{i+1}$ from both branches are then fused through element-wise addition to aggregate contextual information.

PSCCS starts at the deepest feature level and progressively calibrates adjacent features upwards:

$$\begin{aligned}C_1 &= \text{CU}_1(F_1, F_2) \\ C_2 &= \text{CU}_2(C_1, F_3) \\ &\vdots \\ C_{n-1} &= \text{CU}_{n-1}(C_{n-2}, F_n)\end{aligned} \tag{20}$$

The calibrated features are aggregated through simple addition. This progressive calibration strategy effectively resolves spatial misalignment issues between long-range features, enabling SCGC-Net to handle multi-scale features more accurately. This approach enhances the model's performance in landslide detection, particularly for small-scale landslide regions and boundaries, which are highly sensitive to fine spatial details.

IV. Experiments

A. Experimental configuration and evaluation metrics

1) Datasets: This study utilizes four open-source optical remote sensing landslide datasets with pixel-level annotations. These datasets exhibit considerable diversity in terms of geographic environments, landslide types, and imaging conditions, providing a robust basis for a thorough evaluation of the proposed method's performance. Fig. 5 illustrates typical examples of landslide images from these datasets, reflecting the complexity and challenges inherent in landslide detection.

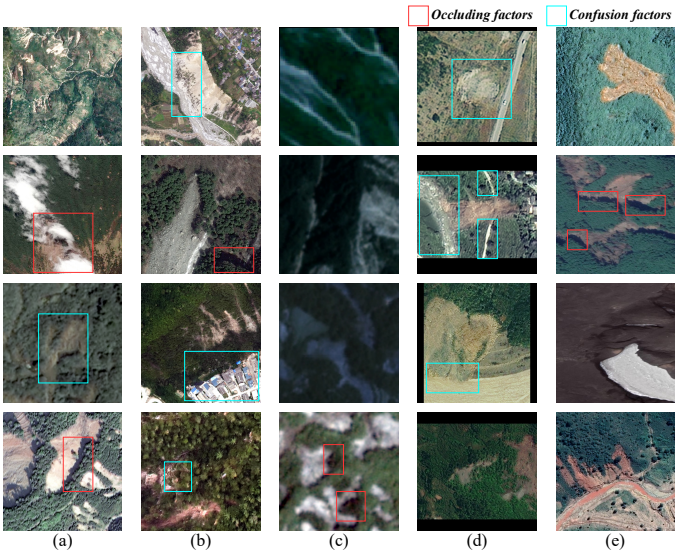

Fig. 5. Landslide datasets. (a) Example of satellite imagery from the CAS Landslide Dataset. (b) Example of UAV imagery from the CAS Landslide Dataset. (c) Example from the HR-GLDD dataset. (d) Example from the Bijie dataset. (e) Example from the GVLM dataset. Red boxes indicate occluded areas, and blue boxes indicate confounding factors.

Among them, the CAS Landslide Dataset [11] is a large-scale, multi-sensor dataset that includes 20,865 images from nine regions, with spatial resolutions ranging from 5m to 0.2m. The HR-GLDD dataset [8] focuses on rainfall- and earthquake-induced landslides in 10 geographic regions worldwide, consisting of 1,758 image tiles from PlanetScope satellites with a resolution of 3m. The Bijie dataset [4] specializes in landslides in the Bijie region of Guizhou, China, comprising 770 before-and-after images from TripleSat satellites at a resolution of 0.8m. The GVLM dataset [9], an independent generalization test set, consists of bi-temporal images from 17 distinct landslide locations captured by Google Earth, with a high resolution of 0.59m, providing a novel and challenging test environment.

Together, these datasets reveal the primary challenges in landslide detection: differences in spatial resolution, the diversity of geographic environments, and various interference factors. In Fig. 5, red boxes highlight occlusion factors, such as clouds and shadows, which directly impact the visibility of landslide areas. Blue boxes indicate confusion factors, such as buildings, roads, and rivers, which visually resemble landslides and may lead to false detections. These factors not only reduce the visibility and boundary clarity of landslides but also significantly complicate detection and accurate identification. To fully assess SCGC-Net's performance, we divided the CAS Landslide Dataset, HR-GLDD, and Bijie datasets into training, validation, and test sets with a 5:3:2 ratio, merging them during training to form a comprehensive training set comprising 11,691/6,949/4,749 images for training, validation, and testing, respectively. The GVLM dataset was reserved for evaluating the model's generalization capability, offering a new and challenging test environment. This diverse dataset configuration ensures the reliability and generalizability of the experimental results, providing a solid foundation for evaluating the real-world applicability of SCGC-Net. By testing the model's performance across varied conditions, we can thoroughly assess the proposed method's effectiveness and adaptability.

2) Evaluation metrics: To thoroughly evaluate the performance of SCGC-Net and other comparative methods in landslide detection, this study employs five widely recognized quantitative metrics: Precision, Recall, F1 Score, IoU, and Overall Accuracy (OA). These metrics are also commonly used in other landslide detection research [28], [53]. Each of these metrics captures a different aspect of model performance and is particularly suited to tasks like landslide detection, which often involve class imbalance and demand precise boundary localization. The metrics are defined as follows:

$$Precision = \frac{TP}{TP + FP}. \tag{21}$$

$$Recall = \frac{TP}{TP + FN}. \tag{22}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{23}$$

$$IoU = \frac{TP}{TP + TN + FP}. \tag{24}$$

$$OA = \frac{TP + TN}{TP + TN + FN + FP}. \tag{25}$$

In these equations, TP (True Positive) refers to landslide pixels correctly identified by the model, FP (False Positive) refers to non-landslide pixels incorrectly classified as landslide pixels, TN (True Negative) represents correctly identified non-landslide pixels, and FN (False Negative) represents landslide pixels that were missed. Precision reflects the model's ability to minimize false positives, while Recall indicates its ability to minimize false negatives. The F1 Score balances Precision and Recall, providing a holistic measure of model performance that is useful for comparing models. IoU measures the overlap between the predicted output and the ground truth, serving as a key indicator of segmentation accuracy. Overall Accuracy assesses the model's performance across the entire dataset.

3) Loss Function: To enhance the accuracy of binary semantic segmentation for landslides, particularly in addressing class imbalance, scale variation, and blurred boundaries, this study proposes a composite loss function that integrates dynamic weighted cross-entropy loss, Focal Loss, and Lovász-Softmax loss. This combination effectively tackles issues related to class imbalance, difficult sample recognition, and boundary precision. The overall loss function is formulated as follows:

$$L = \alpha L_{DWCE}(y, \hat{y}) + \beta L_{Focal}(y, \hat{y}) + \gamma L_{\text{Lovász}}(y, \hat{y}) \tag{26}$$

Here, L_{DWCE} represents the dynamic weighted cross-entropy loss, which adjusts the weights dynamically to account for varying class distributions across different images:

$$L_{DWCE} = -\frac{1}{N}\sum_{i=1}^{N}[w_1(x)\cdot y_i \log(\hat{y}_i) + w_0(x)\cdot(1-y_i)\log(1-\hat{y}_i)] \tag{27}$$

Where N is the total number of pixels, and $w_1(x)$ and $w_0(x)$ are dynamic weights for the landslide and non-landslide classes, respectively, based on the input image x. This adaptive mechanism improves the model's ability to handle diverse landslide distributions.

L_{Focal}, or Focal Loss, helps address class imbalance and enhances the detection of hard-to-recognize samples by modulating the weights of different pixels:

$$L_{\text{Focal}}(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}[y_i(1-\hat{y}_i)^{\gamma}\log(\hat{y}_i) + (1-y_i)\hat{y}_i^{\gamma}\log(1-\hat{y}_i)]. \tag{28}$$

In this equation, γ is the modulation factor, which is set to 2 in this study. Focal Loss improves the detection of small-scale landslides and areas that are difficult to identify.

$L_{\text{Lovász}}$, or Lovász-Softmax loss, is defined as:

$$L_{\text{Lovász}} = \frac{1}{|C|}\sum_{c\in C}\overline{\Delta_{J_c}}(m(c)). \tag{29}$$

The Lovász-Softmax loss is designed to directly optimize the IoU metric, which enhances the overall quality of the segmentation results. In this equation, C is the set of classes, $\overline{\Delta_{J_c}}$ is the Lovász extension, and $m(c)$ is the error vector for class c. The coefficients α, β, and γ are hyperparameters used to balance the relative contributions of each loss component.

4) Implementation Details: All experiments in this study were conducted on an Ubuntu 20.04 system, utilizing an NVIDIA Tesla A800 GPU (80GB) for model training and evaluation. The experiments were implemented using the PyTorch deep learning framework [54]. To maintain consistency across experiments, all input images were resized to 512×512 pixels, with a batch size of 16 for training. Data augmentation techniques included random scale scaling, random cropping (to maintain the 512×512 size), horizontal flipping with a 50% probability, and photometric distortion. For the GVLM dataset, images were first split into smaller tiles, and those without landslides were removed in the generalization experiments to reduce class imbalance. The Adam optimizer was used with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, an initial learning rate of 2e-4, and a weight decay of 0.0001. A cosine annealing schedule was employed to adjust the learning rate. During training, the composite loss function, consisting of L_{DWCE}, L_{Focal}, and $L_{\text{Lovász}}$ (as shown in Equation 26), was used to optimize the model. The weight coefficients α, β, and γ were optimized using a grid search on the validation set. The search space for α, β, and γ was set to the range $[0, 1]$, with a step size of 0.1, and subject to the constraint $\alpha + \beta + \gamma = 1$ to ensure weight normalization. The average F1 score was used as the evaluation metric. The optimal combination of weights was determined to be $\alpha = 0.4$, $\beta = 0.4$, and $\gamma = 0.2$. The total number of training epochs was set to 400.

B. Comparison with State-of-the-Art Methods

To thoroughly evaluate the performance of SCGC-Net, this study compares it with various advanced deep learning methods across three representative datasets. These methods include classical CNN-based architectures, attention-enhanced networks, emerging models based on Transformers and Mamba, as well as hybrid approaches. Specifically, the comparison includes FCN [12], UNet [13], ResUNet [55], PSPNet [44], DeepLab V3+ [42], HRNet [56], ICNet [57], CCNet [58], DANet [59], GCNet [60], SegNext [61], PIDNet [62], SegFormer [36], Swin Transformer [35], ConvNeXt [63], ST-UNet [64], and VMamba [20]. The following sections provide both quantitative and qualitative analyses of SCGC-Net's performance.

Table I presents the performance comparison between SCGC-Net and other advanced methods on the CLD, HR-GLDD, and Bijie datasets. SCGC-Net achieved the best performance across most evaluation metrics, demonstrating its robustness and wide applicability in landslide detection.

On the CLD dataset, SCGC-Net stands out, achieving an IoU of 87.85%, 2.30% improvement over the second-best IoU score. This notable improvement emphasizes SCGC-Net's effectiveness in high-resolution and diverse geographical environments. The increase in IoU indicates that SCGC-Net can more accurately delineate landslide areas and boundary contours, which is critical for assessing landslide extent and risks. The hybrid multi-scale information extraction module plays a crucial role in capturing features across different scales, while the progressive feature calibration fusion strategy effectively integrates these features, enhancing detection precision. For the HR-GLDD dataset, SCGC-Net surpasses the second-best model by 1.88% in Recall and 2.61% in IoU, underscoring its sensitivity in detecting smaller, less noticeable landslides. High recall is essential for landslide risk assessment, as it reduces missed detections and improves the reliability of early warning systems. SCGC-Net's Context-Aware Landslide Modulator plays a key role in adjusting feature weights dynamically, improving the detection of subtle landslide features.

On the Bijie dataset, SCGC-Net achieved an F1 score of 84.75% and an IoU of 73.53%, outperforming the second-best method by 2.08% and 3.06%, respectively. Although VMamba slightly surpasses SCGC-Net in Recall with 85.70%, SCGC-Net compensates with higher precision, ultimately achieving better results in the F1 and IoU metrics. This result indicates SCGC-Net's excellent performance even on smaller datasets, highlighting its strong generalization capabilities. The progressive feature calibration fusion strategy helps mitigate feature misalignment, improving model adaptability across datasets of different scales.

To provide a comprehensive visual comparison of different methods' performance, we selected representative examples from each dataset and compared the detection results from various state-of-the-art approaches, as shown in Fig. 6. SCGC-Net's predictions demonstrate higher integrity and accuracy in various complex scenarios.

In the CLD satellite dataset, SCGC-Net delivers more

TABLE I
COMPARISON RESULTS OF VARIOUS METHODS ON CLD, GLDD, AND BIJIE DATASETS

Method	CLD					GLDD					Bijie				
	P(%)	R(%)	F1(%)	IoU(%)	OA(%)	P(%)	R(%)	F1(%)	IoU(%)	OA(%)	P(%)	R(%)	F1(%)	IoU(%)	OA(%)
FCN [12]	88.64	82.45	85.43	74.57	95.89	78.85	64.11	70.72	54.70	91.23	82.73	69.13	75.32	60.41	95.31
U-Net [13]	88.09	83.01	85.47	74.63	95.93	78.51	67.69	72.70	57.11	94.45	65.12	69.13	67.07	50.45	94.12
ResUNet [55]	84.09	86.83	85.44	74.58	95.74	80.16	74.99	77.49	63.25	95.24	81.91	71.44	76.32	61.71	96.16
PSP-Net [44]	90.39	88.52	89.44	80.90	96.26	78.65	71.88	75.11	60.15	94.80	81.51	72.50	76.74	62.26	96.19
Deeplabv3+ [42]	91.38	87.22	89.26	80.60	96.97	80.07	72.92	76.33	61.72	95.06	82.17	71.40	76.41	61.82	96.18
HRNet [56]	92.05	87.33	89.63	81.20	97.09	83.51	72.46	77.59	63.38	95.43	80.84	78.68	79.75	66.31	96.54
ICNet [57]	78.18	88.11	82.85	70.72	94.74	69.22	75.42	72.19	56.48	93.65	84.06	64.10	72.73	57.15	95.84
CCNet [58]	92.04	89.31	90.68	82.96	97.17	78.14	72.51	75.22	60.28	94.78	78.08	72.59	75.24	60.30	95.86
GCNet [60]	91.59	88.45	90.06	81.80	96.98	82.10	76.64	79.28	65.67	95.62	79.11	75.69	77.36	63.08	96.16
PIDNet [62]	88.58	83.43	85.93	75.32	96.06	79.90	67.65	73.27	57.81	94.61	79.95	75.92	77.88	63.78	96.27
DANet [59]	92.15	89.17	90.64	82.88	97.34	83.72	78.20	80.86	67.88	95.96	80.70	71.39	75.76	60.98	96.04
ConvNeXt [63]	93.22	91.14	92.17	85.47	97.75	85.08	78.79	81.81	69.23	96.21	81.14	84.26	82.67	70.47	96.94
SegFormer [36]	92.29	87.48	89.82	81.52	97.14	82.31	76.44	79.27	65.66	95.63	76.76	79.94	78.32	64.36	96.17
SwinTrans [35]	93.62	90.85	92.21	85.55	97.81	84.22	78.78	81.41	68.65	96.08	82.72	80.20	81.44	68.69	96.83
ST-UNet [64]	91.62	90.60	91.11	83.67	97.45	82.99	79.53	81.22	68.38	96.02	81.83	82.48	82.16	69.72	96.90
VMamba [20]	92.04	90.67	91.35	84.08	97.53	83.28	80.22	81.72	69.09	96.17	79.79	85.70	82.64	70.41	96.88
Ours	94.75	92.35	93.53	87.85	98.16	85.18	82.10	83.61	71.84	96.49	85.04	84.46	84.75	73.53	97.37

comprehensive detection of small-scale landslides, effectively reducing false positives from bare soil, shallow vegetation, clouds, and shadows. This is largely due to the hybrid multi-scale information extraction module, which captures landslide features at multiple scales.

For the CLD UAV dataset, SCGC-Net excels in detecting large-scale landslides, maintaining detection integrity even under partial vegetation cover. Additionally, SCGC-Net shows strong discrimination in complex scenes involving intersecting roads, rivers, and landslides, due to the progressive feature calibration fusion strategy that effectively integrates multi-scale features, enhancing the model's understanding of complex terrains.

On the HR-GLDD dataset, SCGC-Net excels at detecting low-contrast, high-confusion landslides, correlating with the significant increase in Recall. The context-aware landslide modulator plays a key role in dynamically adjusting feature weights to enhance sensitivity to weak landslide features.

Finally, results from the Bijie dataset further confirm SCGC-Net's generalization capabilities. Even on a small dataset, SCGC-Net delivers optimal prediction performance, showcasing its adaptability. This advantage stems from SCGC-Net's architecture, especially its progressive feature calibration fusion strategy, which mitigates feature bias introduced by different dataset scales. In summary, SCGC-Net demonstrates significant advantages across various datasets and landslide types. It excels in detecting small targets and irregularly shaped landslides while reducing missed detections and improving edge clarity. SCGC-Net's strong discrimination in complex environments reduces false positives and enhances overall detection accuracy. These superior results stem from SCGC-Net's unique design, including hybrid multi-scale information extraction, progressive feature calibration fusion, and an Context-Aware Landslide Modulator, enabling the model to better adapt to the challenges presented by different datasets.

C. Ablation Experiment

To evaluate the effectiveness of the individual components of SCGC-Net and validate the rationality of its network structure, we conducted comprehensive ablation experiments on the CLD dataset, focusing on module ablation and the influence of dilation rates.

TABLE II
ABLATION STUDY ON SCGC-NET MODULES

Exp ID	HMIE	CAM	PSCCS	P (%)	R (%)	F1 (%)	IoU (%)	OA (%)
A1				89.77	87.01	88.37	79.16	96.15
A2	✓			92.66	89.77	91.20	83.80	97.32
A3		✓		90.05	89.08	89.60	81.10	96.27
A4			✓	90.74	88.61	89.66	81.26	96.34
A5	✓	✓		93.11	91.86	92.48	86.01	97.75
A6	✓		✓	92.75	90.48	91.60	84.50	97.62
A7		✓	✓	92.20	92.17	92.18	85.50	97.24
A8	✓	✓	✓	94.75	92.35	93.53	87.85	98.16

In the module ablation experiment, we sequentially introduced the HMIE, CAM, and PSCCS modules to quantify their contributions to model performance. Table II presents the quantitative results for different module combinations. The baseline model (Experiment A1), which uses ResNet50 as the backbone network and incorporates a feature pyramid structure for feature fusion, achieved an IoU of 79.16

Introducing the HMIE module (Experiment A2) led to a significant performance improvement, raising the IoU to 83.80%, an increase of 4.64%. This improvement can be attributed to the HMIE module's ability to combine Mamba and CNN architectures effectively, enhancing the model's capacity to capture multi-scale features and long-range dependencies. The impact of the HMIE module is even more pronounced when combined with other modules. For instance, the combination of HMIE and CAM (Experiment A5) boosted the IoU to 86.01%, while the combination of HMIE and PSCCS (Experiment A6) resulted in an IoU of 84.50%. These results demonstrate the efficacy of HMIE in extracting complex landslide features and its compatibility with other modules. When used in isolation, the CAM module (Experiment A3) improved the recall by 2.07% over the baseline model. Although the improvement was smaller compared to HMIE, CAM exhibited strong synergistic effects when combined with other modules. Notably, the combination of CAM and PSCCS (Experiment

Fig. 6. Landslide detection results comparing with 16 state-of-the-art approaches across different datasets (CLD Satellite, CLD UAV, HR-GLDD, and Bijie), including traditional CNN-based methods, lightweight model, attention-enhanced methods, modern backbone network, transformer-based models, mamba-based model. (a) Optical RSIs; (b) Ground Truth(GT); (c) FCN; (d) U-Net; (e) ResUNet; (f) PSPNet; (g) DeepLabv3+; (h) HRNet; (i) ICNet; (j) CCNet; (k) GCNet; (l) PIDNet; (m) DANet; (n) ConvNeXt; (o) SegFormer; (p) Swin Transformer; (q) ST-UNet; (r) VMamba; (s) SCGC-Net(ours).

A7) yielded outstanding recall performance, reaching 92.17%, the highest recall rate apart from the complete model. This suggests that the CAM module, through its context-awareness mechanism, significantly enhances the model's ability to comprehend complex landslide scenes, particularly in cases of blurred boundaries and challenging backgrounds. The PSCCS module alone (Experiment A4) increased the IoU to 81.26%. While the effect of PSCCS in isolation was relatively modest, it played a crucial role when combined with other modules. For example, combining PSCCS with HMIE (Experiment A6) increased the IoU to 84.50%, further improving performance compared to using HMIE alone. This highlights the effectiveness of PSCCS in addressing spatial alignment issues during feature fusion, making a significant contribution to improving the accuracy of landslide boundary localization.

The complete SCGC-Net model (Experiment A8) achieved the best performance across all evaluation metrics, with an IoU of 87.85%, representing an 8.69% improvement over the baseline model. This significant performance boost underscores the synergistic effect of the three core modules: HMIE provides robust multi-scale feature extraction, CAM refines semantic information through its context-awareness mechanism, and PSCCS effectively resolves spatial alignment challenges in feature fusion.

To visually demonstrate the influence of SCGC-Net's core modules on feature extraction, we performed a feature map

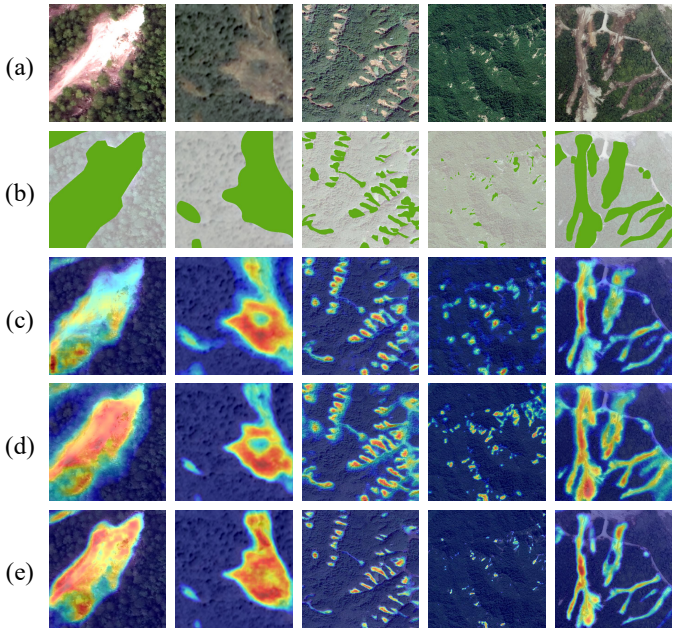

Fig. 7. Feature map visualization from the ablation experiment. (a) Optical RSIs; (b) Ground truth; (c) Feature map from Experiment A2; (d) Feature map from Experiment A5; (e) Feature map from Experiment A8.

TABLE III
ABLATION STUDY ON ATROUS RATES

Exp ID	SCFE1	SCFE2	P (%)	R (%)	F1 (%)	IoU (%)
B1	$R_{1,1,1}$	$R_{1,1,1}$	92.77	89.01	90.85	83.23
B2	$R_{2,2,2}$	$R_{5,5,5}$	93.59	90.45	91.99	85.17
B3	$R_{2,2,2}$+$R_{2,2,2}$	$R_{3,3,3}$+$R_{3,3,3}$	94.12	92.16	93.13	87.14
B4	$R_{3,3,3}$+$R_{3,3,3}$	$R_{5,5,5}$+$R_{5,5,5}$	92.99	92.50	92.74	86.47
B5	$R_{3,3,3}$+$R_{5,5,5}$	$R_{8,8,8}$+$R_{12,12,12}$	93.38	88.53	90.89	83.30
B6	$R_{2,2,2}$+$R_{3,3,3}$	$R_{3,3,3}$+$R_{5,5,5}$	94.75	92.35	93.53	87.85
B7	$R_{1,1,1}$+$R_{2,2,2}$	$R_{1,2,2}$+$R_{2,2,3}$	94.75	92.35	93.53	87.85

visualization analysis on key module combinations from the ablation experiment, as shown in Fig. 7. The feature map generated by the HMIE module alone (Fig. 7c) already highlights the landslide regions, confirming HMIE's effectiveness in capturing multi-scale features. With the introduction of the CAM module (Fig. 7d), the feature map presents more refined semantic information, displaying clearer landslide boundaries and richer internal structures. This improvement is particularly evident in small landslides and complex terrains, illustrating CAM's enhancement of the model's semantic refinement capabilities. The feature map produced by the complete SCGC-Net model (Fig. 7e) shows the most accurate representation of landslides, with well-defined boundaries and detailed internal structures. This outcome visually demonstrates the contribution of the PSCCS module in aligning multi-scale features, consistent with the quantitative results. The ablation experiment results not only validate the effectiveness of each component within SCGC-Net but also reveal the synergistic interactions between them. By systematically combining and analyzing these modules, we gained a deeper understanding of their specific contributions to improving model performance, providing crucial experimental evidence and theoretical support for further optimization of landslide detection algorithms.

The dilation rate of convolutions is a critical parameter that influences the size of the receptive field. Its primary role is to mitigate the limitations of the Mamba structure in local feature extraction, while balancing the extraction of fine-grained local details and broader contextual information. To further explore the effect of different dilation rate combinations on model performance, we conducted an ablation experiment on dilation rates within the SCFE Block of the HMIE module in SCGC-Net. Table III provides a detailed comparison of the model's performance under various dilation rate configurations.

The results indicate that the introduction of multi-scale dilation rates significantly enhances the model's performance. Taking Experiment B1 as the baseline, which employs standard convolutions with a dilation rate of 1, the model effectively captures fine local features but struggles with complex landslide patterns. In contrast, Experiments B2 through B7, which incorporate various combinations of dilation rates, demonstrate superior multi-scale feature extraction, both locally and regionally. The progressive small dilation rate strategy adopted in Experiments B6 and B7 yielded the best results, achieving the highest F1 scores and IoU values. This strategy involves the use of different dilation rate combinations across the two SCFE modules, allowing the network to capture features at multiple scales. This design enables the model to attend to receptive fields of varying sizes and spatial details, significantly improving its ability to detect multi-scale landslide features. However, Experiment B5 shows that using excessively large dilation rates leads to performance degradation, with the IoU decreasing to 83.30%. This decline may be attributed to the disruption of local spatial continuity caused by large dilation rates, resulting in the loss of detail and breakdown of contextual information, which impairs the effective extraction and integration of multi-scale features.

These results underscore the importance of carefully balancing local feature extraction and regional contextual information when designing the MSLE Block within the SCFE module. The progressive small dilation rate strategy, which gradually increases dilation rates across different SCFE modules, ensures that the model remains sensitive to fine spatial details while simultaneously expanding the receptive field to enhance the capture of regional spatial features. This approach not only addresses the limitations of the Mamba structure in local feature extraction but also retains its strength in capturing long-range dependencies.

D. Generalization Analysis

To assess SCGC-Net's generalization performance when confronted with new data domains, we designed an incremental learning experiment using the GVLM dataset. The GVLM dataset includes large-scale landslide images and corresponding mask data from various geographic regions worldwide, with a spatial resolution of 0.59m. Since the data source (Google Earth) and resolution differ from the training datasets, it provides an ideal test environment to evaluate the model's cross-domain generalization capabilities.

In this experiment, we selected several high-performing models from previous experiments, including DANet, ConvNeXt, Swin Transformer, VMamba, and our proposed SCGC-Net. First, all models were trained on the CAS Landslide

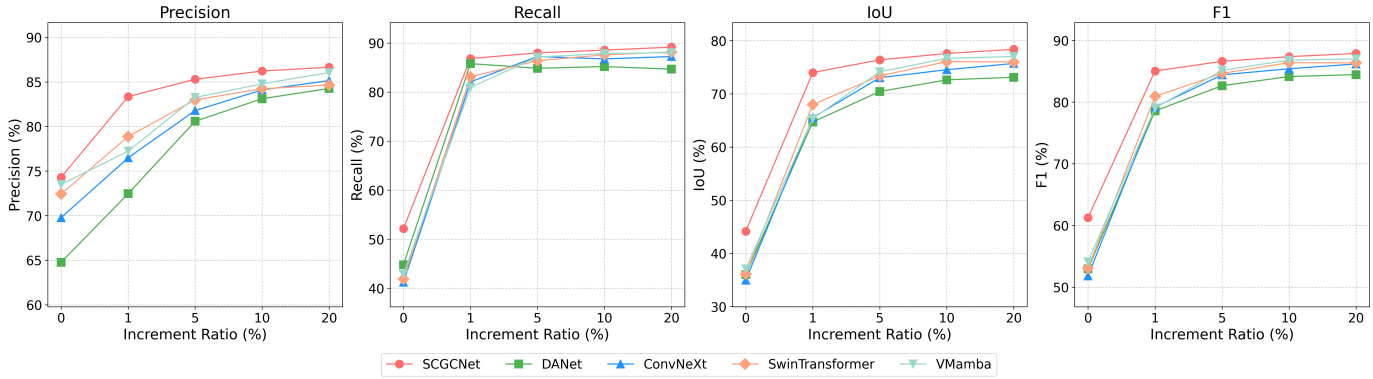

Fig. 8. Comparison of model accuracy with different percentages of domain incremental data on the GVLM dataset.

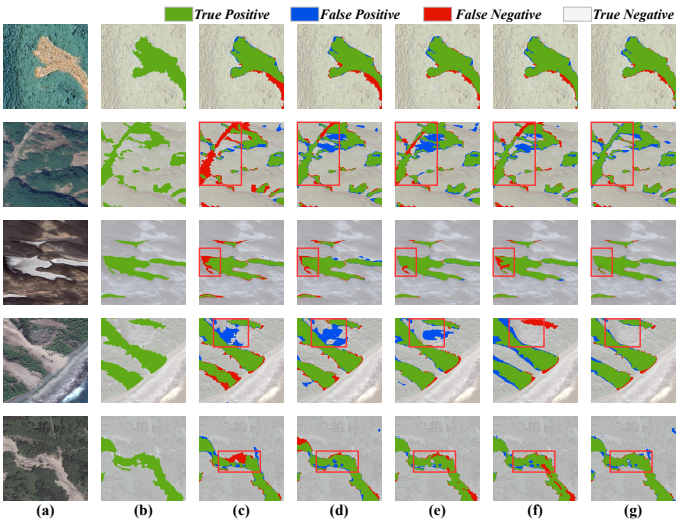

Fig. 9. Detection results after 5% incremental fine-tuning on the GVLM dataset, comparing with 4 representative state-of-the-art models. (a) Optical RSIs; (b) Ground truth; (c) DANet; (d) ConvNeXt; (e) Swin Transformer; (f) VMamba; (g) SCGC-Net(ours).

Dataset (CLD), HR-GLDD, and Bijie datasets. Then, the GVLM dataset was divided into 512×512 image patches, and 20% of the patches containing landslides and 20% of those without landslides were randomly selected as domain incremental data. We defined incremental percentages of 0% (direct testing on GVLM), 1%, 5%, 10%, and 20%, and for each percentage, a corresponding amount of GVLM incremental data was randomly added to the training set for fine-tuning. This setup simulates real-world scenarios where a model needs to adapt to new and unseen data domains. Finally, we tested the models on the remaining GVLM data, using Precision, Recall, F1-score, and IoU as evaluation metrics.

Fig. 8 shows the performance variation of each model under different percentages of GVLM incremental data. The results indicate that, with 0% increment (direct transfer), SCGC-Net significantly outperforms other models in terms of Recall, IoU, and F1 metrics, demonstrating its superior cross-dataset generalization performance. As the percentage of GVLM data increases, all models improve, but SCGC-Net adapts the fastest. With only 1% of GVLM data, SCGC-Net's performance stabilizes, whereas other models require 5% or more data to achieve comparable results. Notably, SCGC-Net maintains its performance lead at all incremental levels. Even at the 20% increment, SCGC-Net's IoU and F1 metrics remain about 1% higher than the second-best model. As the incremental ratio reaches 20%, the performance of all models begins to plateau, suggesting that the models have adequately adapted to the GVLM dataset's feature distribution.

To further assess the models' generalization ability, we conducted a qualitative analysis. Fig. 9 shows the landslide detection visual results on the GVLM dataset after 5% incremental fine-tuning for each model. We highlight regions with significant performance differences among models using red boxes. In simpler landslide scenarios, all models are able to accurately identify landslide boundaries. However, in more complex scenarios, such as irregularly shaped landslides and densely packed small-scale landslides (e.g., rows two, four, and five in Fig. 9), SCGC-Net performs the best, producing more complete landslide extractions with sharper boundaries and effectively reducing false positives. In regions with greater spectral variation (e.g., row three in Fig. 9), although all models exhibit good adaptability after fine-tuning, SCGC-Net still achieves the highest accuracy in feature extraction. These observations further confirm SCGC-Net's superiority in handling complex and diverse landslide scenarios.

The experimental results clearly demonstrate SCGC-Net's exceptional generalization capacity to quickly adapt to new data domains. This ability is crucial in practical applications, as it allows SCGC-Net to rapidly adjust to new landslide detection environments with minimal additional data, greatly reducing the cost of data collection and annotation. SCGC-Net's superior generalization performance can be attributed to several key factors: first, the Spatial Context Guide Calibration mechanism effectively captures and leverages spatial contextual information, enabling the model to adapt more easily to variations in terrain and landslide characteristics across different datasets. Second, SCGC-Net's multi-scale contextual feature extraction and fine-grained fusion strategies allow the model to simultaneously focus on local details and global structures, enhancing its robustness. Finally, the adaptive nature of the CAM enables the model to better adjust its focus to match the feature distribution in new datasets.

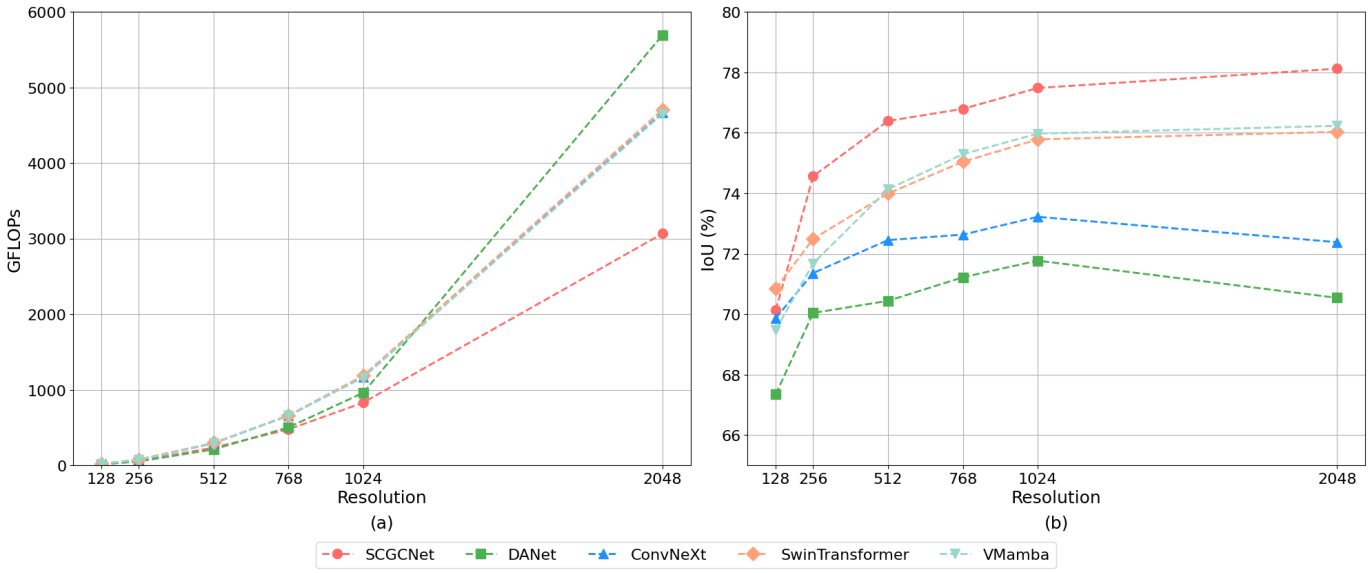

Fig. 10. Comparative analysis of model efficiency and accuracy at different image resolutions

E. Computational Efficiency and Performance Analysis

In large-scale landslide detection tasks, processing large remote sensing images is often necessary to capture sufficient geospatial context. Therefore, the model's detection accuracy and computational efficiency across different image sizes are critical for practical applications. To evaluate this, we compared the landslide detection performance of DANet, ConvNeXt, Swin Transformer, VMamba, and SCGC-Net using remote sensing images of varying sizes. All models were fine-tuned with a 10% data increment to ensure fairness. The experiment tested six different input resolutions: 128×128, 256×256, 512×512, 768×768, 1024×1024, and 2048×2048. Computational efficiency was measured in GFLOPs (billion floating point operations), and detection accuracy was evaluated using IoU.

Fig. 10 illustrates the comparative performance of each model across different input resolutions. The results show that SCGC-Net clearly outperforms other models in computational efficiency for larger image sizes ($\geq$1024×1024). Notably, at the 2048×2048 resolution, SCGC-Net requires only 3066 GFLOPs, which is significantly lower than other models (ranging from 4600 to 5700). The relatively slow increase in computational complexity with larger image sizes highlights SCGC-Net's excellent scalability, which is crucial for large-scale landslide detection tasks.

Regarding detection accuracy, SCGC-Net consistently achieves the highest IoU across all resolutions, improving from 70.13% at 128×128 to 78.12% at 2048×2048, marking the most significant improvement. This indicates that SCGC-Net can effectively leverage the detailed information and spatial context provided by high-resolution images, which is essential for accurately detecting landslide boundaries in complex terrains. Other models, such as Swin Transformer and VMamba, also show performance improvements, but due to their similar architectures, ConvNeXt, Swin Transformer, and VMamba exhibit similar quadratic increases in computational complexity as image size grows. Notably, DANet shows not only a decline in detection performance with larger images but also a steep quadratic increase in computational cost. This is primarily due to the global attention mechanism, which scales poorly as image size increases, combined with the inherent limitation of CNNs' local receptive fields. In contrast, SCGC-Net's hybrid architecture strikes a strong balance between scalability and performance. At the 2048×2048 resolution, SCGC-Net requires 30%–46% fewer GFLOPs than other models while maintaining the highest IoU, clearly demonstrating its potential for large-scale landslide monitoring applications.

V. Discussion

A. Advantages and Contributions

The comprehensive experimental results presented in Section IV demonstrate the effectiveness of SCGC-Net in multi-source landslide detection tasks through its novel architectural design and effective feature learning strategy. The integration of Mamba and CNN architectures enables comprehensive feature extraction at both global and local scales. This is evidenced by the ablation experiments in Section IV-C, where the complete SCGC-Net achieves an 8.69% IoU improvement over the baseline model. Such performance gain stems from the synergistic effect of three key components: the HMIE module captures multi-scale features efficiently, the CAM module refines semantic information adaptively, and the PSCCS module resolves spatial misalignment issues effectively.

Cross-domain generalization represents a critical capability for practical landslide detection systems. Deep learning models often struggle with domain shifts due to varying geographical conditions and imaging platforms. Our generalization analysis in Section IV-D reveals that SCGC-Net requires only 1% of GVLM data for successful domain adaptation, while competing methods need at least 5% for comparable performance. This superior adaptability can be attributed to two factors:

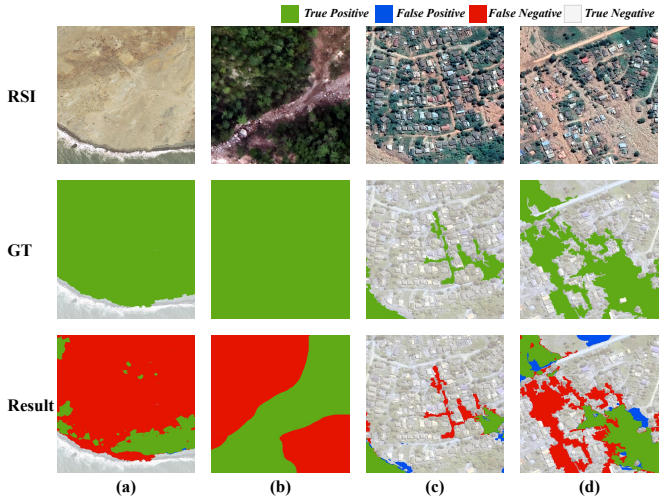

Fig. 11. Examples of detection challenges in high-resolution imagery. (a)-(b) large-scale landslide detection failures; (c)-(d) detection errors in densely built areas.

first, the context-guided calibration mechanism learns domain-invariant features effectively; second, the progressive spatial calibration strategy maintains consistent feature representation across domains. The visual results in Fig. 9 further validate this conclusion, showing robust detection performance across diverse environmental conditions.

Computational efficiency is essential for processing high-resolution remote sensing images in real-time disaster monitoring. Traditional transformer-based approaches often suffer from quadratic computational complexity, limiting their practical application. Our efficiency analysis in Section IV-E demonstrates that SCGC-Net processes 2048×2048 resolution images with only 3066 GFLOPs, reducing computational cost by 30-46% compared to state-of-the-art methods. This improvement results from the linear complexity of Mamba-based feature extraction and the lightweight implementation of the CAM module. The analysis in Fig. 10 confirms SCGC-Net's efficient scaling with image size while maintaining detection accuracy.

B. Limitations and Future Perspectives

By analyzing cases with low IoU scores, we identified two challenges in landslide detection in high-resolution images. First, as shown in Fig. 11(a) and (b), when landslide regions occupy a large portion of the image and individual image patches can only cover a local area of the entire landslide, the detection performance of the model may be compromised, leading to false-negative detection results. Second, as depicted in Fig. 11(c) and 11(d), the coexistence of landslide regions and dense urban structures poses additional challenges, leading to both false negative and false positive detections. These detection issues partly stem from the limitations of the dataset's processing method based on small-sized image patches (512×512 pixels). The limited size of the image patches restricts the model's ability to capture comprehensive contextual information, resulting in semantic discontinuities between adjacent image patches and incomplete landslide detection across multiple image patches. Moreover, the complex interactions between landslide areas and urban infrastructure, combined with insufficient contextual information, affect the model's ability to accurately identify landslide boundaries in urban settings.

To address the above detection challenges, future research will focus on two aspects. First, to address the limitations in large-scale landslide detection and the semantic discontinuities caused by patch-based processing, efficient landslide detection methods based on full-image processing need to be explored. These methods would enable the model to process complete high-resolution remote sensing images while maintaining semantic consistency, thereby reducing the loss of contextual information caused by image patch processing. Second, strategies for integrating multi-source remote sensing data will be explored, such as incorporating LiDAR point clouds and surface classification data. These complementary data sources, which provide three-dimensional terrain and surface information, hold the potential to enhance detection accuracy in complex surface environments, particularly for landslides in urban areas that are currently challenging to identify with precision.

VI. Conclusion

In this study, a novel SCGC-Net for landslide detection in multi-source remote sensing images is proposed, which adopts an asymmetric two-branch feature extraction network combined with Mamba and CNN to efficiently extract global context features and local spatial features. It introduces the Spatial Calibration Fusion Module, which significantly improves the boundary delineation accuracy and small-scale landslide segmentation capability. In addition, a Context-Aware Landslide Modulator is developed to dynamically integrate multi-scale contextual information. Experimental results on three representative datasets demonstrate that SCGC-Net exhibits the best performance under different environmental conditions and imaging platforms. Moreover, SCGC-Net demonstrates excellent generalization ability in cross-domain experiments based on the GVLM dataset.

Future research can explore the potential application of this architecture in other geohazard detection tasks, and further investigate its performance and robustness in multimodal data fusion for more comprehensive hazard detection and assessment.

References

[1] U. Ozturk, E. Bozzolan, E. A. Holcombe, R. Shukla, F. Pianosi, and T. Wagener, "How climate change and unplanned urban sprawl bring more landslides," *Nature*, vol. 608, no. 7922, pp. 262–265, 2022.

[2] A. Stumpf and N. Kerle, "Object-oriented mapping of landslides using random forests," *Remote sensing of environment*, vol. 115, no. 10, pp. 2564–2577, 2011.

[3] Z. Lv, T. Liu, X. Kong, C. Shi, and J. A. Benediktsson, "Landslide inventory mapping with bitemporal aerial remote sensing images based on the dual-path fully convolutional network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4575–4584, 2020.

[4] S. Ji, D. Yu, C. Shen, W. Li, and Q. Xu, "Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks," *Landslides*, vol. 17, pp. 1337–1352, 2020.

[5] Y. Liu and L. Wu, "Geological disaster recognition on optical remote sensing images using deep learning," *Procedia Computer Science*, vol. 91, pp. 566–575, 2016.

[6] A. Novellino, C. Pennington, K. Leeming, S. Taylor, I. G. Alvarez, E. McAllister, C. Arnhardt, and A. Winson, "Mapping landslides from space: A review," *Landslides*, pp. 1–12, 2024.

[7] T. Lei, Y. Zhang, Z. Lv, S. Li, S. Liu, and A. K. Nandi, "Landslide inventory mapping from bitemporal images using deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 982–986, 2019.

[8] S. R. Meena, L. Nava, K. Bhuyan, S. Puliero, L. P. Soares, H. C. Dias, M. Floris, and F. Catani, "Hr-gldd: A globally distributed dataset using generalized dl for rapid landslide mapping on hr satellite imagery," *Earth System Science Data Discussions*, vol. 2022, pp. 1–21, 2022.

[9] X. Zhang, W. Yu, M.-O. Pun, and W. Shi, "Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 197, pp. 1–17, 2023.

[10] C. Zeng, Z. Cao, F. Su, Z. Zeng, and C. Yu, "High-precision aerial imagery and interpretation dataset of landslide and debris flow disaster in sichuan and surrounding areas," *China Sci. Data*, vol. 7, no. 2, pp. 195–205, 2022.

[11] Y. Xu, C. Ouyang, Q. Xu, D. Wang, B. Zhao, and Y. Luo, "Cas landslide dataset: A large-scale and multisensor dataset for deep learning-based landslide detection," *Scientific Data*, vol. 11, no. 1, p. 12, 2024.

[12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[14] Y. Yi and W. Zhang, "A new deep-learning-based approach for earthquake-triggered landslide detection from single-temporal rapideye satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6166–6176, 2020.

[15] W. Xia, J. Chen, J. Liu, C. Ma, and W. Liu, "Landslide extraction from high-resolution remote sensing imagery using fully convolutional spectral–topographic fusion network," *Remote Sensing*, vol. 13, no. 24, p. 5116, 2021.

[16] Z. Dong, S. An, J. Zhang, J. Yu, J. Li, and D. Xu, "L-unet: A landslide extraction model using multi-scale feature fusion and attention mechanism," *Remote Sensing*, vol. 14, no. 11, p. 2552, 2022.

[17] P. Lv, L. Ma, Q. Li, and F. Du, "Shapeformer: A shape-enhanced vision transformer model for optical remote sensing image landslide detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 2681–2689, 2023.

[18] Y. Huang, J. Zhang, H. He, Y. Jia, R. Chen, Y. Ge, Z. Ming, L. Zhang, and H. Li, "Mast: An earthquake triggered landslides extraction method combining morphological analysis edge recognition with swin-transformer deep learning model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

[19] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[20] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, and Y. Liu, "Vmamba: Visual state space model," 2024. [Online]. Available: https://arxiv.org/abs/2401.10166

[21] X. Ma, X. Zhang, and M.-O. Pun, "Rs 3 mamba: Visual state space model for remote sensing image semantic segmentation," *IEEE Geoscience and Remote Sensing Letters*, 2024.

[22] J. Ruan and S. Xiang, "Vm-unet: Vision mamba unet for medical image segmentation," *arXiv preprint arXiv:2402.02491*, 2024.

[23] H. Cai, T. Chen, R. Niu, and A. Plaza, "Landslide detection using densely connected convolutional networks and environmental conditions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5235–5247, 2021.

[24] O. Ghorbanzadeh, Y. Xu, P. Ghamisi, M. Kopp, and D. Kreil, "Landslide4sense: Reference benchmark data and deep learning models for landslide detection," *arXiv preprint arXiv:2206.00515*, 2022.

[25] B. Yu, C. Xu, F. Chen, N. Wang, and L. Wang, "Hadeennet: A hierarchical-attention multi-scale deconvolution network for landslide detection," *International Journal of Applied Earth Observation and Geoinformation*, vol. 111, p. 102853, 2022.

[26] W. Lu, Y. Hu, W. Shao, H. Wang, Z. Zhang, and M. Wang, "A multiscale feature fusion enhanced cnn with the multiscale channel attention mechanism for efficient landslide detection (ms2landsnet) using medium-resolution remote sensing data," *International Journal of Digital Earth*, vol. 17, no. 1, p. 2300731, 2024.

[27] X. Zheng, L. Han, G. He, N. Wang, G. Wang, and L. Feng, "Semantic segmentation model for wide-area coseismic landslide extraction based on embedded multichannel spectral–topographic feature fusion: A case study of the jiuzhaigou ms7. 0 earthquake in sichuan, china," *Remote Sensing*, vol. 15, no. 4, p. 1084, 2023.

[28] K. Wang, D. He, Q. Sun, L. Yi, X. Yuan, and Y. Wang, "A novel network for semantic segmentation of landslide areas in remote sensing images with multi-branch and multi-scale fusion," *Applied Soft Computing*, vol. 158, p. 111542, 2024.

[29] X. Liu, Y. Peng, Z. Lu, W. Li, J. Yu, D. Ge, and W. Xiang, "Feature-fusion segmentation network for landslide detection using high-resolution remote sensing images and digital elevation model data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–14, 2023.

[30] H. Li, Y. He, Q. Xu, J. Deng, W. Li, and Y. Wei, "Detection and segmentation of loess landslides via satellite images: A two-phase framework," *Landslides*, vol. 19, no. 3, pp. 673–686, 2022.

[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[32] Z. Lu, Y. Peng, W. Li, J. Yu, D. Ge, L. Han, and W. Xiang, "An iterative classification and semantic segmentation network for old landslide detection using high-resolution remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[33] R. Fu, J. He, G. Liu, W. Li, J. Mao, M. He, and Y. Lin, "Fast seismic landslide detection based on improved mask r-cnn," *Remote Sensing*, vol. 14, no. 16, p. 3928, 2022.

[34] O. Ghorbanzadeh, Y. Xu, H. Zhao, J. Wang, Y. Zhong, D. Zhao, Q. Zang, S. Wang, F. Zhang, Y. Shi *et al.*, "The outcome of the 2022 landslide4sense competition: Advanced landslide detection from multisource satellite imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 9927–9942, 2022.

[35] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[36] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in neural information processing systems*, vol. 34, pp. 12 077–12 090, 2021.

[37] P. Li, Y. Wang, T. Si, K. Ullah, W. Han, and L. Wang, "Mffsp: Multi-scale feature fusion scene parsing network for landslides detection based on high-resolution satellite images," *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107337, 2024.

[38] L. Wu, R. Liu, N. Ju, A. Zhang, J. Gou, G. He, and Y. Lei, "Landslide mapping based on a hybrid cnn-transformer network and deep transfer learning using remote sensing images with topographic and spectral features," *International Journal of Applied Earth Observation and Geoinformation*, vol. 126, p. 103612, 2024.

[39] H. Chen, Y. He, L. Zhang, S. Yao, W. Yang, Y. Fang, Y. Liu, and B. Gao, "A landslide extraction method of channel attention mechanism u-net network based on sentinel-2a remote sensing images," *International Journal of Digital Earth*, vol. 16, no. 1, pp. 552–577, 2023.

[40] O. Ghorbanzadeh, K. Gholamnia, and P. Ghamisi, "The application of resu-net and obia for landslide detection from multi-temporal sentinel-2 images," *Big Earth Data*, vol. 7, no. 4, pp. 961–985, 2023.

[41] R. Wei, C. Ye, T. Sui, H. Zhang, Y. Ge, and Y. Li, "A feature enhancement framework for landslide detection," *International Journal of Applied Earth Observation and Geoinformation*, vol. 124, p. 103521, 2023.

[42] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[43] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[44] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.

[45] X. Wang, D. Wang, T. Sun, J. Dong, L. Xu, W. Li, S. Li, P. Ran, J. Ao, Y. Zou *et al.*, "Dual path attention network (dpanet) for intelligent

identification of wenchuan landslides," *Remote Sensing*, vol. 15, no. 21, p. 5213, 2023.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[47] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.
[48] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," *arXiv preprint arXiv:2401.09417*, 2024.
[49] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural networks*, vol. 107, pp. 3–11, 2018.
[50] J. Yang, C. Li, X. Dai, and J. Gao, "Focal modulation networks," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4203–4217, 2022.
[51] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
[52] Z. Huang, Y. Wei, X. Wang, W. Liu, T. S. Huang, and H. Shi, "Alignseg: Feature-aligned segmentation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 550–557, 2021.
[53] Y. Xie, N. Zhan, J. Zhu, B. Xu, H. Chen, W. Mao, X. Luo, and Y. Hu, "Landslide extraction from aerial imagery considering context association characteristics," *International Journal of Applied Earth Observation and Geoinformation*, vol. 131, p. 103950, 2024.
[54] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.
[55] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.
[56] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5693–5703.
[57] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, "Icnet for real-time semantic segmentation on high-resolution images," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 405–420.
[58] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 603–612.
[59] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3146–3154.
[60] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "Gcnet: Non-local networks meet squeeze-excitation networks and beyond," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019.
[61] M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "Segnext: Rethinking convolutional attention design for semantic segmentation," *Advances in Neural Information Processing Systems*, vol. 35, pp. 1140–1156, 2022.
[62] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired by pid controllers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 19 529–19 539.
[63] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 976–11 986.
[64] X. He, Y. Zhou, J. Zhao, D. Zhang, R. Yao, and Y. Xue, "Swin transformer embedding unet for remote sensing image semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

Yukun Fan received the B.S. degree from the School of Geographic Information Science, China University of Geosciences, Beijing, China, in 2022. He is currently pursuing the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include disaster monitoring, computer vision, and multisource data fusion for remote sensing applications.

Peifeng Ma (Senior Member, IEEE) received the M.S. degree in signal and information processing from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2012, and the Ph.D. degree in earth system and geoinformation science from the Institute of Space and Earth Information Science, The Chinese University of Hong Kong (CUHK), Hong Kong, in 2016. He is currently an Assistant Professor with the Department of Geography and Resource Management, Institute of Space and Earth Information Science, CUHK. His research interests include persistent scatterer interferometry, synthetic aperture radar tomography, distributed scatterer interferometry.

Qingbo Hu received the M.S. degree in railway engineering from Southwest Jiaotong University, Chengdu, China, in 1991. He is currently a Senior Engineer with China Railway Economic and Planning Research Institute Co. Ltd., Beijing, China. His research interests include geological remote sensing technology and its applications in railway engineering, with particular focus on geological hazard assessment and monitoring.

Guiwei Liu received the Ph.D. degree from the Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, in 2010. He is currently a Senior Engineer with China Railway Design Corporation, Tianjin, China. His research interests include geological remote sensing, artificial intelligence, and disaster assessment technology and their applications in railway engineering and infrastructure development.

Zihuan Guo received the B.S. degree in electronic information engineering from the Hefei University of Technology, China, in 2021. He is currently pursuing the Ph.D. degree with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include polarimetric SAR and disaster analysis.

Yixian Tang received the Ph.D. degree in cartography and geographic information systems from the Institute of Remote Sensing Applications, Chinese Academy of Sciences(CAS), Beijing, China, in 2006. He is currently an Associate Professor with the Aerospace Information Research Institute, CAS, and the International Research Center of Big Data for Sustainable Development Goals, Beijing, China. His research interests include differential SAR interferometry, time series InSAR methodology, and their applications in surface displacement monitoring, with particular expertise in processing large-scale InSAR data for regional deformation analysis.

Fan Wu (Member, IEEE) received the master's degree in surveying and mapping engineering from Wuhan University, Wuhan, China, in 2002, and the Ph.D. degree in microwave remote sensing from the Institute of Remote Sensing Application, Chinese Academy of Sciences (CAS), Beijing, China, in 2005. He is currently an Associate Professor with the International Research Center of Big Data for Sustainable Development Goals, Aerospace Information Research Institute, CAS. His research interests include SAR image processing, information extraction techniques, and their applications in urban change detection and disaster analysis.

Hong Zhang (Member, IEEE) received the B.S. and the M.S. degrees from Beijing Normal University, Beijing, China, in 1990 and 1994, respectively, and the Ph.D. degree from Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 2002. Now she is a professor at the International Research Center of Big Data for Sustainable Development Goals (CBAS), Aerospace Information Research Institute, CAS, Beijing, China. Her research interests include polarimetric SAR methods and applications, interferometric SAR applications and so on.